

45131
SEARCH REQUEST FORM

Requestor's Name: CHRIS DRABIK Serial Number: 09/445201

Date: 6/19/01 Phone: 605 1156 Art Unit: 1633
1234

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

PLEASE SEARCH SEPARATELY THE FOLLOWING
SEGMENTS OF SEQ ID #1 (NUCLEIC ACID ONLY)

A 8260 → 10560

B 8336 → 10608

C 10094 → 10608
ED-

D 6036 → 6959

IF 100% HITS SHOW UP FOR A, B, & C (i.e. ALL 3
MUST BE 100%)
PLEASE USE THOSE HITS TO SEARCH AGAINST
THE ENTIRE SEQ ID #1 SEQUENCE.

THANKS

- CHRIS

9 HRS

STAFF USE ONLY

Date completed: 6/20/01
6/22/01
Searcher: _____
Terminal time: _____
Elapsed time: _____
CPU time: _____
Total time: _____
Number of Searches: _____
Number of Databases: _____

Search Site
☒ STIC
☒ CM-1

Type of Search
☒ A.A. Sequence
☒ A.A. Sequence

Bibliographic

Vendors

IG

STN

Dialog

APS

Geninfo

SDC

DARC/Questel

Other

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 17:58:27 ; Search time 346.98 Seconds
(without alignments)
3824.240 Million cell updates/sec

Title: US-09-445-201-1_COPY_8336_10608

Perfect score: 2273

Sequence: 1 tctcagacagacgcgtgtt.....aggcttctcttgatgcc 2273

Scoring table: IDENTITY_NUC

Gapex 10.0, Gapex 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_0401.*
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2273	100.0	12845	20 V84274	Mouse endothelial
2	55.4	2.4	5513	21 C63529	Slit protein codin
3	46	2.0	4064	21 A07587	Mouse piwi gene, d
4	45.6	2.0	51259	18 X83007	Partial mouse WRN
5	45	2.0	2362	21 Z50462	Human tissue trans
6	44	1.9	1110	19 V29343	Calcium ion channe
7	43.6	1.9	810	20 Z16495	Human gene express
8	43.2	1.9	11901	20 X02998	Human IL-1ra BAC c
9	42.6	1.9	8966	20 Z09581	Human Apo A1 genom
10	42.2	1.9	14507	18 T73568	Expression augment
11	42.2	1.9	14507	21 Z88869	Chinese hamster 2A

c 12 42 1.8 687 20 X37433 Human secreted pro
c 13 833 1.8 833 21 C98178 Human colon cancer
c 14 41.8 1.8 1751 19 V59610 Human secreted pro
c 15 41.8 1.8 41684 21 A28150 Human purH gene ge
c 16 41.6 1.8 10120 20 Z23683 Human DRCl DNA fra
c 17 41.6 1.8 119950 20 X02021 Human yeast gene.
c 18 41.4 1.8 32351 21 F21307 Human low adenosin
c 19 41.4 1.8 32351 21 A35185 Human low adenosin
c 20 41.4 1.8 40298 21 F21311 Human low adenosin
c 21 41.4 1.8 40298 21 A35189 Human adenosine re
c 22 41.2 1.8 1502 21 C59687 Human secreted pro
c 23 41 1.8 84607 20 X90847 Human PACAP genomi
c 24 40.8 1.8 837 20 X37525 Human secreted pro
c 25 40.8 1.8 1863 20 X85010 Human secreted pro
c 26 40.8 1.8 4116 20 X02983 Human secreted pro
c 27 40.8 1.8 4736 20 X23315 Human IL-1ra BAC c
c 28 40.6 1.8 4034 20 X35011 Mouse IL-alpha-OHas
c 29 40.6 1.8 6511 16 Q95493 Human glucose 6-ph
c 30 40.4 1.8 209 16 T26063 Human gene signatu
c 31 40.4 1.8 267 21 C29151 Human secreted pro
c 32 40.4 1.8 1838 20 X27332 Human secreted pro
c 33 40.4 1.8 4696 20 X03041 Human secreted pro
c 34 40.4 1.8 1035 21 C60001 Human IL-1ra BAC c
c 35 40.2 1.8 6901 18 T78853 Human secreted pro
c 36 40.2 1.8 49999 20 Z23891 Human lecithin-cho
c 37 40.2 1.8 49999 20 Z23891 Murine LOBO genomi
c 38 40.2 1.8 49999 20 Z23896 Murine LOBO homolo
c 39 40.2 1.8 138169 21 A34791 Human adenosine re
c 40 40.2 1.8 141589 21 F20913 Human ELAM-1 polyn
c 41 40.2 1.8 141589 21 F21127 Human ELAM-1 polyn
c 42 40.2 1.8 141589 21 F21152 Human low adenosin
c 43 40.2 1.8 141589 21 A35005 Human low adenosin
c 44 40.2 1.8 141589 21 A35030 Human adenosine re
c 45 40.2 1.8 146981 21 F21442 Human ELAM-1 polyn

ALIGNMENTS

RESULT 1
V84274
ID V84274 standard; DNA; 12845 BP.
XX AC V84274;
XX DT 12-APR-1999 (first entry)
XX DE Mouse endothelial growth factor receptor-2 gene Flk-1 5' region.
XX FK-1; endothelial growth factor receptor-2; VEGF; transcription;
KW endothelium; enhancer; vascular disease; angiogenesis; cancer;
KW diabetic retinopathy; rheumatoid arthritis; wound healing;
KW vulnary; atherosclerosis; tumour; neuronal disorder; therapy;
KW diagnosis; mouse; ss.
XX OS Mus sp.
XX FH Location/Qualifiers
FT Promoter 6036..6959
FT Exon /*tag= a
6661..7026
/*tag= b
FT Intron /number= 1
7027..10632
/*tag= c
/*tag= d
/note= "DNA sequences comprising nucleotides
8260-10560, 8336-10608 and/or 10094-10608
are specifically claimed in Claim 3b"
protein_bind 10137..10155
/*tag= d
/note= "GATA/PEA3 consensus binding site"

CC cells preferably nerve axons or dendrites, leukocytes or malignant cells
 CC expressing roundabout (robo) protein, by repulsion or inhibition. Also,
 CC the slit protein is useful for treating graft rejection, inhibiting the
 CC infection of a cell by HIV, inhibiting inflammation and in aiding wound
 CC repair and organ regeneration. In addition, the slit protein is useful
 CC for the treatment of conditions involving the migration of leukocytes,
 CC such as asthma, arthritis, glomerulonephritis, cystic fibrosis,
 CC ulcerative colitis, Crohn's disease, multiple sclerosis, allergic
 CC encephalomyelitis, Alzheimer's disease and coronary artery restenosis.
 XX
 SQ Sequence 5513 BP; 1668 A; 1216 C; 1211 G; 1418 T; 0 other;

Query Match 2.4%; Score 55.4; DB 21; Length 5513;
 Best Local Similarity 57.1%; Pred. No. 3.6e-05;
 Matches 101; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
 QY 1593 ctccagcttgatggtatgctcagagctttgattaaatgagtgagacccctagct 1652
 DB 4114 CTGCAGAGTGATGTGGCTGAGCGGGATCTTTTCGAGGAATGGGATTAACGAGTC 4055
 QY 1653 attgctcattagactacactatttttagttttgctctgagtttatgaatgcatgtat 1712
 DB 4054 TTTGTTTCACAAATTTAGCGTGAAGTATTTTCACATTTGTTCCACGATAGCCAGGCAG 3995
 QY 1713 gcatgaactggagagatatttttctcccaattctcttctcatttaaatgtgc 1769
 DB 3994 GCACGGCACAGAAGATCATTCCTCTTTCACAAATGATGGGCTCCATTTTCACATTC 3938

RESULT 3

A07587/c
 ID A07587 standard; DNA; 4064 BP.

AC A07587;
 AC A07587;
 DT 29-AUG-2000 (first entry)
 DE Mouse piwi gene, designated miwi.
 XX
 KW Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
 KW anaemia; immunodeficiency; male infertility; mouse; ds.
 XX
 OS Mus sp.

Key Location/Qualifiers
 CDS 191..2779
 FT /*tag= a
 FT /product= miwi
 FT /transl_except= (pos:450..452; aa:Xaa)
 FT /transl_except= (pos:1337..1339; aa:Xaa)
 FT /transl_except= (pos:2636..2638; aa:Xaa)
 FT /note= "Xaa= Leu or Ile"
 FT /transl_except= (pos:836..838; aa:Xaa)
 FT /note= "Xaa= unspecified amino acid"

WO200032039-A1.

08-JUN-2000.

03-DEC-1999; 99WO-US28764.

04-DEC-1998; 98US-0110901.

(UYDU-) UNIV DUKE.

Lin H;

WPI; 2000-412085/35.

P-PSDB; Y90234.

XX Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
 XX therapy of diseases such as cancer and in various research and
 PT

PT diagnostic applications -

XX Claim 19; Page 180-185; 201pp; English.

XX This sequence encodes the mouse piwi family protein, designated
 CC miwi. The piwi family nucleic acids and polypeptides are used in gene
 CC therapy of diseases such as cancer and also in various research and
 CC diagnostic applications. The sequences can also be used to treat
 CC tissue dystrophy, anaemia, immunodeficiency, and male infertility.
 XX
 SQ Sequence 4064 BP; 1114 A; 978 C; 1077 G; 890 T; 5 other;

Query Match 2.0%; Score 46; DB 21; Length 4064;
 Best Local Similarity 63.6%; Pred. No. 0.014;
 Matches 70; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 1298 ttatgtattctctctctctgagacagagatctcaactgtgtgcccagctagctctcaaaact 1357
 DB 3656 TTCTTTTCTCTCTCTCTGTTTAAAGACAAAGTCTCACTATGTAGCCTAGGCTGACCGGAAC 3597
 QY 1358 tgcgggtccattttgtctcaactcaccagaatgctgggcttcagggtgtgtgc 1407
 DB 3596 TGGGGTTCCTCGGCTCACCTCCTGAGTCTGAGATCACAGGCGGTGTC 3547

RESULT 4

ID X83007 standard; DNA; 51259 BP.

XX X83007;
 AC X83007;

DT 31-AUG-1999 (first entry)
 DE Partial mouse WRN genomic sequence #3.

XX Mouse; WRN; Werner's syndrome; detection: diagnosis; autosomal;
 KW recessive disorder; phenotype; ss.
 XX
 OS Mus musculus.

PN WO9724435-A1.

PD 10-JUL-1997.

PF 30-DEC-1996; 96WO-US20785.

PR 12-APR-1996; 96US-0632175.

PR 29-DEC-1995; 95US-0009409.

PR 29-DEC-1995; 95US-0580539.

PR 30-JAN-1996; 96US-0010835.

PR 30-JAN-1996; 96US-0594242.

XX (DARW-) DARWIN MOLECULAR CORP.
 PA (OSHI/) OSHIMA J.

PI Fu Y, Mulligan J, Oshima J, Schellenberg GB, Yu C;
 WPI; 1997-363671/33.

XX Isolated nucleic acid molecule encoding the WRN gene product -

PT useful for detection and treatment of Werner's syndrome, and related
 PT diseases

XX Claim 1; Fig 7; 153pp; English.

XX This sequence represents a fragment of the genomic sequence containing
 CC the coding region for the mouse WRN gene (X83004). The corresponding
 CC human gene (X83001) encodes a protein related to Werner's syndrome.
 CC The products can be used for the detection and treatment of Werner's
 CC syndrome (WS), an autosomal recessive disorder with a complex phenotype,
 CC as well as related diseases.

XX


```

FT      /*tag= h
FT      /note= "the EASE sequence is preferably ligated
FT      to DNA comprising nucleotides 14290-14507
FT      (Claim 3),"
XX      WO9725420-A1.
XX      PN
XX      PD 17-JUL-1997.
XX      XX
XX      PF 10-JAN-1997; 97WO-US00483.
XX      PR 11-JAN-1996; 96US-0586509.
XX      PA (IMMV ) IMMUNEX CORP.
XX      XX
XX      PI Lee C, Morris AE, Thomas JN;
XX      WI; 1997-372861/34.
XX      XX
XX      PT Expression augmenting sequence elements - used in vectors for
XX      PT expressing recombinant proteins at high levels in shorter periods of
XX      PT time
XX      PS Claim 1; Page 23-30; 36pp; English.
XX      CC A 14507 bp DNA sequence (T73568) comprises a novel transcription
XX      CC regulatory element, expression augmenting sequence element (EASE),
XX      CC that facilitates high expression of recombinant proteins in
XX      CC mammalian host cells. It was identified by cloning the integration
XX      CC site of a unique expression cassette encoding recombinant dimeric
XX      CC tumour necrosis factor receptor-immunoglobulin Fc fusion protein
XX      CC from genomic DNA of a CHO 2A5-3 cell line expressing this protein
XX      CC at a high level. Expression vectors incorporating an EASE show
XX      CC a 2- to 8-fold improvement of recombinant protein expression levels
XX      CC in CHO host cells. The EASE sequence also facilitates high
XX      CC expression in shorter periods of time.
XX      XX
XX      SQ Sequence 14507 BP; 4225 A; 2548 C; 3136 G; 4598 T; 0 other;

Query Match 1.9%; Score 42.2; DB 18; Length 14507;
Best Local Similarity 59.9%; Pred. No. 0.32;
Matches 88; Conservative 0; Mismatches 58; Indels 1; Gaps 1;

QY 1288 ctgctattaattagttattctctctgtgagacagagtcactgtgtgcccaggcta 1347
Db 146 cagccaagatttgttgtttgtttccgagaaaggttctctgtgagccctgtg 205

QY 1348 gtctcaact-tgcggtccattgtctcactcactcagaatgctgggcttcagggtgtg 1406
Db 206 tcagggaactcagagatccaccgtctctctctgagtgctgggattaaagggtgtg 265

QY 1407 caccacactaggtagctgcggttttaa 1433
Db 266 ctgactacaggaagctgtgttttta 292

RESULT 11
ID Z88869 standard; DNA; 14507 BP.
AC Z88869;
XX Z88869;
DT 25-MAY-2000 (first entry)
XX Chinese hamster 2A5-3 lambda EASE DNA.
DE EASE; expression augmenting sequence element; chinese hamster; ds.
XX Crictetus sp.
XX US6027915-A.
XX

```

```

PD 22-FEB-2000.
XX 13-JAN-1997; 97US-0785150.
XX 11-JAN-1996; 96US-0586509.
XX (IMMV ) IMMUNEX CORP.
XX Morris AE, Thomas JN, Lee C;
XX WI; 2000-194831/17.
XX Improved expression of recombinant proteins in stable cell pools of
XX mammalian cells comprises transforming cells with a vector containing
XX an expression augmenting sequence element -
XX Claim 1; Column 17-30; 18pp; English.
XX This invention describes a novel expression augmenting sequence element
XX (EASE) (I) isolated from Chinese hamster ovary cells. (I) facilitates
XX high expression of recombinant proteins in mammalian host cells in a
XX short time period. (I) is used for improving expression of recombinant
XX proteins 2-8 fold in stable cell pools when it is present in an
XX expression vector. (I) improves recombinant protein expression in
XX mammalian cells and facilitates the use of mammalian cells in
XX recombinant protein production.
XX Sequence 14507 BP; 4225 A; 2548 C; 3136 G; 4598 T; 0 other;

Query Match 1.9%; Score 42.2; DB 21; Length 14507;
Best Local Similarity 59.9%; Pred. No. 0.32;
Matches 88; Conservative 0; Mismatches 58; Indels 1; Gaps 1;

QY 1288 ctgctattaattagttattctctctgtgagacagagtcactgtgtgcccaggcta 1347
Db 146 cagccaagatttgttgtttgtttccgagaaaggttctctgtgagccctgtg 205

QY 1348 gtctcaact-tgcggtccattgtctcactcactcagaatgctgggcttcagggtgtg 1406
Db 206 tcagggaactcagagatccaccgtctctctctgagtgctgggattaaagggtgtg 265

QY 1407 caccacactaggtagctgcggttttaa 1433
Db 266 ctgactacaggaagctgtgttttta 292

RESULT 12
X37433/C
ID X37433 standard; cDNA; 687 BP.
XX X37433;
AC X37433;
XX 06-JUL-1999 (first entry)
XX Human secreted protein cDNA fragment containing gene 65.
XX Human; secreted protein; prevention; treatment; protein therapy;
XX gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
XX developmental abnormality; foetal deficiency; blood disorder; lymphoma;
XX leukemia; immune system disorder; autoimmune disease; hepatic disease;
XX renal disease; inflammation; allergy; asthma; sepsis; diabetes; AIDS;
XX Alzheimer's disease; cognitive disorder; schizophrenia; osteoporosis;
XX arthritis; psoriasis; digestive; endocrine; infection; ss.
XX Homo sapiens.
XX WO9909155-A1.
XX 25-FEB-1999.
XX 18-AUG-1998; 98WO-US17044.
XX

```


XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX Homo sapiens.

XX WO9839448-A2.

XX 11-SEP-1998.

XX 06-MAR-1998; 98WO-US04493.

XX 03-OCT-1997; 97US-0061060.

XX 07-MAR-1997; 97US-0038621.

XX 07-MAR-1997; 97US-0040161.

XX 07-MAR-1997; 97US-0040162.

XX 07-MAR-1997; 97US-0040163.

XX 07-MAR-1997; 97US-0040333.

XX 07-MAR-1997; 97US-0040334.

XX 07-MAR-1997; 97US-0040336.

XX 07-MAR-1997; 97US-0040626.

XX 11-APR-1997; 97US-0043311.

XX 11-APR-1997; 97US-0043312.

XX 11-APR-1997; 97US-0043313.

XX 11-APR-1997; 97US-0043314.

XX 11-APR-1997; 97US-0043568.

XX 11-APR-1997; 97US-0043569.

XX 11-APR-1997; 97US-0043576.

XX 11-APR-1997; 97US-0043578.

XX 11-APR-1997; 97US-0043580.

XX 11-APR-1997; 97US-0043669.

XX 11-APR-1997; 97US-0043670.

XX 11-APR-1997; 97US-0043671.

XX 11-APR-1997; 97US-0043672.

XX 23-MAY-1997; 97US-0043674.

XX 23-MAY-1997; 97US-0047492.

XX 23-MAY-1997; 97US-0047500.

XX 23-MAY-1997; 97US-0047501.

XX 23-MAY-1997; 97US-0047502.

XX 23-MAY-1997; 97US-0047503.

XX 23-MAY-1997; 97US-0047581.

XX 23-MAY-1997; 97US-0047582.

XX 23-MAY-1997; 97US-0047583.

XX 23-MAY-1997; 97US-0047584.

PR 23-MAY-1997; 97US-0047633.

PR 06-JUN-1997; 97US-0048964.

PR 13-JUN-1997; 97US-0048974.

PR 08-JUL-1997; 97US-0049610.

PR 16-JUL-1997; 97US-0051926.

PR 18-AUG-1997; 97US-0052874.

PR 22-AUG-1997; 97US-0055724.

PR 22-AUG-1997; 97US-0056630.

PR 22-AUG-1997; 97US-0056631.

PR 22-AUG-1997; 97US-0056632.

PR 22-AUG-1997; 97US-0056636.

PR 22-AUG-1997; 97US-0056637.

PR 22-AUG-1997; 97US-0056662.

PR 22-AUG-1997; 97US-0056664.

PR 22-AUG-1997; 97US-0056845.

PR 22-AUG-1997; 97US-0056862.

PR 22-AUG-1997; 97US-0056864.

PR 22-AUG-1997; 97US-0056872.

PR 22-AUG-1997; 97US-0056874.

PR 22-AUG-1997; 97US-0056875.

PR 22-AUG-1997; 97US-0056876.

PR 22-AUG-1997; 97US-0056877.

PR 22-AUG-1997; 97US-0056878.

PR 22-AUG-1997; 97US-0056879.

PR 22-AUG-1997; 97US-0056880.

PR 22-AUG-1997; 97US-0056881.

PR 22-AUG-1997; 97US-0056882.

PR 22-AUG-1997; 97US-0056884.

PR 22-AUG-1997; 97US-0056886.

PR 22-AUG-1997; 97US-0056887.

PR 22-AUG-1997; 97US-0056888.

PR 22-AUG-1997; 97US-0056889.

PR 22-AUG-1997; 97US-0056892.

PR 22-AUG-1997; 97US-0056893.

PR 22-AUG-1997; 97US-0056894.

PR 22-AUG-1997; 97US-0056903.

PR 22-AUG-1997; 97US-0056908.

PR 22-AUG-1997; 97US-0056909.

PR 22-AUG-1997; 97US-0056910.

PR 05-SEP-1997; 97US-0056911.

PR 05-SEP-1997; 97US-0057650.

PR 05-SEP-1997; 97US-0057669.

PR 12-SEP-1997; 97US-0057761.

XX 97US-0058785.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;

XX Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS;

XX Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;

XX Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;

XX WPI; 1998-506364/43.

XX P-PSDB; W74830.

XX New isolated human genes and the secreted polypeptide(s) they encode

XX - useful for diagnosis and treatment of e.g. cancers, neurological

XX disorders, immune diseases, inflammation or blood disorders

XX Claim 1; Page 333-334; 721pp; English.

XX This sequence represents a nucleic acid molecule designated Gene 101 from

XX the human cDNA clone HLQAB52 (deposited as clone ATCC 97901 and ATCC

XX 209047) which encodes a secreted human protein. The gene can be used to

XX generate fusion proteins by linking to the gene to a human immunoglobulin

XX Fc portion (e.g. V59502) for increasing the stability of the fused

XX protein as compared to the human protein only.

XX The invention relates to 186 novel genes and their fragments (nucleic

XX acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which

XX are useful for preventing, treating or ameliorating medical conditions

XX e.g. by protein or gene therapy. Also, pathological conditions can be

XX diagnosed by determining the amount of the new polypeptides in a sample

XX or by determining the presence of mutations in the new polynucleotides.

CC Specific uses are described for each of the 186 polynucleotides, based on
 CC which tissues they are most highly expressed in (see V59511 for described
 CC uses).

XX Sequence 1751 BP; 405 A; 477 C; 476 G; 393 T; 0 other;

Query Match 1.8%; Score 41.8; DB 19; Length 1751;
 Best Local Similarity 61.5%; Pred. No. 0.13;
 Matches 67; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Oy 1292 tattaattagttattcttcttgagacagagtcactcgtgtgcccaggctagctc 1351

Db 1750 TTTTTTTTTTTTTTTTTTTTGGACGGAGTCTGCTCTGCGCCAGGCTGGACT 1691

Oy 1352 caaactgcggtccattgtctcactcatcagaatgctgggtccagg 1400

Db 1690 GCGGACTGACGTGGCGGCAATCTCGGCTCACTGCAAGCTCCGCTCCCGG 1642

RESULT 15

A28150
 ID A28150 standard; DNA; 41684 BP.

XX A28150;

XX 29-JAN-2001 (first entry)

XX Human purH gene genomic DNA sequence.

XX Human; purH; biallelic marker; single nucleotide polymorphism; SNP;
 KW diagnosis; prevention; treatment; prostate cancer; tumour;
 KW chromosome 1q34-q35; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

FT 5'UTR 1..2000

FT exon /*tag= a

FT exon /*tag= b

FT intron /*tag= c

FT exon /*tag= d

FT intron /*tag= e

FT exon /*tag= f

FT intron /*tag= g

FT exon /*tag= h

FT intron /*tag= i

FT exon /*tag= j

FT intron /*tag= k

FT exon /*tag= l

FT intron /*tag= m

FT exon /*tag= n

FT intron /*tag= o

FT exon /*tag= p

FT intron /*tag= q

FT exon /*tag= r

FT intron /*tag= s

FT exon /*tag= t

FT intron /*tag= u

FT exon /*tag= v

FT intron /*tag= w

FT exon /*tag= x

FT intron /*tag= y

FT exon /*tag= z

FT intron /*tag= aa

FT exon /*tag= ab

FT intron /*tag= ac

FT exon /*tag= ad

FT intron /*tag= ae

FT exon /*tag= af

FT intron /*tag= ag

FT exon /*tag= ah

FT intron /*tag= ai

FT exon /*tag= aj

FT intron /*tag= ak

FT exon /*tag= al

FT intron /*tag= am

FT exon /*tag= an

FT intron /*tag= ao

FT exon /*tag= ap

FT intron /*tag= aq

FT exon /*tag= ar

FT intron /*tag= as

FT exon /*tag= at

FT intron /*tag= au

FT exon /*tag= av

FT intron /*tag= aw

FT exon /*tag= ax

FT intron /*tag= ay

FT exon /*tag= az

FT intron /*tag= ba

FT exon /*tag= bb

FT intron /*tag= bc

FT exon /*tag= bd

FT intron /*tag= be

FT exon /*tag= bf

FT intron /*tag= bg

FT exon /*tag= bh

FT intron /*tag= bi

FT exon /*tag= bj

FT intron /*tag= bk

FT exon /*tag= bl

FT intron /*tag= bm

FT exon /*tag= bn

FT intron /*tag= bo

FT exon /*tag= bp

FT intron /*tag= bq

FT exon /*tag= br

FT intron /*tag= bs

FT exon /*tag= bt

FT intron /*tag= bu

FT exon /*tag= bv

FT intron /*tag= bw

FT exon /*tag= bx

FT intron /*tag= by

FT exon /*tag= bz

FT intron /*tag= ca

FT exon /*tag= cb

FT intron /*tag= cc

FT exon /*tag= cd

FT intron /*tag= ce

FT exon /*tag= cf

FT intron /*tag= cg

FT exon /*tag= ch

FT intron /*tag= ci

FT exon /*tag= cj

FT intron /*tag= ck

FT exon /*tag= cl

FT intron /*tag= cm

FT exon /*tag= cn

FT intron /*tag= co

FT exon /*tag= cp

FT intron /*tag= cq

FT exon /*tag= cr

FT intron /*tag= cs

FT exon /*tag= ct

FT intron /*tag= cu

FT exon /*tag= cv

FT intron /*tag= cw

FT exon /*tag= cx

FT intron /*tag= cy

FT exon /*tag= cz

FT intron /*tag= da

FT exon /*tag= db

FT intron /*tag= dc

FT exon /*tag= dd

FT intron /*tag= de

FT exon /*tag= df

FT intron /*tag= dg

FT exon /*tag= dh

FT intron /*tag= di

FT exon /*tag= dj

FT intron /*tag= dk

FT exon /*tag= dl

FT intron /*tag= dm

FT exon /*tag= dn

FT intron /*tag= do

FT exon /*tag= dp

FT intron /*tag= dq

FT exon /*tag= dr

FT intron /*tag= ds

FT exon /*tag= dt

FT intron /*tag= du

FT exon /*tag= dv

FT intron /*tag= dw

FT exon /*tag= dx

FT intron /*tag= dy

FT exon /*tag= dz

FT intron /*tag= ea

FT exon /*tag= eb

FT intron /*tag= ec

FT exon /*tag= ed

FT intron /*tag= ee

FT exon /*tag= ef

FT intron /*tag= eg

FT exon /*tag= eh

FT intron /*tag= ei

FT exon /*tag= ej

FT intron /*tag= ek

FT exon /*tag= el

FT intron /*tag= em

FT exon /*tag= en

FT intron /*tag= eo

FT exon /*tag= ep

FT intron /*tag= eq

FT exon /*tag= er

FT intron /*tag= es

FT exon /*tag= et

FT intron /*tag= eu

FT exon /*tag= ev

FT intron /*tag= ew

FT exon /*tag= ex

FT intron /*tag= ey

FT exon /*tag= ez

FT intron /*tag= fa

FT exon /*tag= fb

FT intron /*tag= fc

FT exon /*tag= fd

FT intron /*tag= fe

FT exon /*tag= ff

FT intron /*tag= fg

FT exon /*tag= fh

FT intron /*tag= fi

FT exon /*tag= fj

FT intron /*tag= fk

FT exon /*tag= fl

FT intron /*tag= fm

FT exon /*tag= fn

FT intron /*tag= fo

FT exon /*tag= fp

FT intron /*tag= fq

FT exon /*tag= fr

FT intron /*tag= fs

FT exon /*tag= ft

FT intron /*tag= fu

FT exon /*tag= fv

FT intron /*tag= fw

FT exon /*tag= fx

FT intron /*tag= fy

FT exon /*tag= fz

FT intron /*tag= ga

FT exon /*tag= gb

FT intron /*tag= gc

FT exon /*tag= gd

FT intron /*tag= ge

FT exon /*tag= gf

FT intron /*tag= gg

FT exon /*tag= gh

FT intron /*tag= gi

FT exon /*tag= gj

FT intron /*tag= gk

FT exon /*tag= gl

FT intron /*tag= gm

FT exon /*tag= gn

FT intron /*tag= go

FT exon /*tag= gp

FT intron /*tag= gq

FT exon /*tag= gr

FT intron /*tag= gs

FT exon /*tag= gt

FT intron /*tag= gu

FT exon /*tag= gv

FT intron /*tag= gw

FT exon /*tag= gx

FT intron /*tag= gy

FT exon /*tag= gz

FT intron /*tag= ha

FT exon /*tag= hb

FT intron /*tag= hc

FT exon /*tag= hd

FT intron /*tag= he

FT exon /*tag= hf

FT intron /*tag= hg

FT exon /*tag= hh

FT intron /*tag= hi

FT exon /*tag= hj

FT intron /*tag= hk

FT exon /*tag= hl

FT intron /*tag= hm

FT exon /*tag= hn

FT intron /*tag= ho

FT exon /*tag= hp

FT intron /*tag= hq

FT exon /*tag= hr

FT intron /*tag= hs

FT exon /*tag= ht

FT intron /*tag= hu

FT exon /*tag= hv

FT intron /*tag= hw

FT exon /*tag= hx

FT intron /*tag= hy

FT exon /*tag= hz

FT intron /*tag= ia

FT exon /*tag= ib

FT intron /*tag= ic

FT exon /*tag= id

FT intron /*tag= ie

FT exon /*tag= if

FT intron /*tag= ig

FT exon /*tag= ih

FT intron /*tag= ii

FT exon /*tag= ij

FT intron /*tag= ik

FT exon /*tag= il

FT intron /*tag= im

```
FT variation replace(23175,T)
FT /*tag= an
FT replace(23253,G)
FT /*tag= ao
FT variation replace(26106,C)
FT /*tag= ap
FT variation replace(30464,G)
FT /*tag= aq
FT variation replace(30669,C)
FT /*tag= ar
FT variation replace(31250,A)
FT /*tag= as
FT variation replace(35148,G)
FT /*tag= at
FT variation replace(36801,G)
FT /*tag= au
FT variation replace(37286,T)
FT /*tag= av
FT variation replace(37536,C)
FT /*tag= aw
FT variation replace(39321,A)
FT /*tag= ax
FT variation replace(39689,T)
FT /*tag= ay
FT primer_bind 6137..6157
FT /*tag= az
FT primer_bind complement (6577..6597)
FT /*tag= ba
FT primer_bind 14864..14882
FT /*tag= bb
FT primer_bind complement (15292..15312)
FT /*tag= bc
FT primer_bind 15837..15855
FT /*tag= bd
FT primer_bind complement (16429..16266)
FT /*tag= be
FT primer_bind 16599..16617
FT /*tag= bf
FT primer_bind complement (17030..17049)
FT /*tag= bg
FT primer_bind 18131..18150
FT /*tag= bh
FT primer_bind complement (18592..18610)
FT /*tag= bi
FT primer_bind 22710..22727
FT /*tag= bj
FT primer_bind 23100..23118
FT /*tag= bk
FT primer_bind complement (23130..23149)
FT /*tag= bl
FT primer_bind complement (23512..23530)
FT /*tag= bm
FT primer_bind 25822..25840
FT /*tag= bn
FT primer_bind complement (26222..26241)
FT /*tag= bo
FT primer_bind 30332..30352
FT /*tag= bp
FT primer_bind complement (30754..30773)
FT /*tag= bq
FT primer_bind 30918..30935
FT /*tag= br
FT primer_bind complement (31390..31408)
FT /*tag= bs
FT primer_bind 34780..34799
FT /*tag= bt
FT primer_bind complement (35215..35233)
FT /*tag= bu
FT primer_bind 36593..36610
FT /*tag= bv
FT primer_bind complement (37017..37036)
FT /*tag= bw
FT primer_bind 37060..37080
FT /*tag= bx
FT complement (37541..37561)
FT /*tag= by

Query Match 1.8%; Score 41.8; DB 21; Length 41684;
Best Local Similarity 66.1%; Pred. No. 0.72; Indels 2; Gaps 1;
Matches 76; Conservative 0; Mismatches 37;

QY 1295 taattagttattcttcttctgagacagagttctcactgtgtgcccaggctagtctcaa 1354
||||| || ||| ||||||||| |||| ||| ||||||| |||||
Db 4598 taattttttgtatttttggtagagacagagtttcaccatgttgccaggctgactcga 4657
||||| || ||| ||||||||| |||| ||| ||||||| |||||

QY 1355 act--tgcgggtccatttgtctcactcactcagaatgtgggtctccagggtgtgtgc 1407
||||| || ||| || ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4658 actcctgacctcaagtgatcgccgccttgccctgctgggattacagggtgtgagc 4712
||||| || ||| || ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: June 21, 2001, 18:00:40
Job time: 13448 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 18:10:18 ; Search time 8015.43 Seconds
(without alignments)
4386.314 Million cell updates/sec

Title: US-09-445-201-1_COPY_8336_10608
Perfect score: 2273
Sequence: 1 tctcgagacagagcgctgt.....agggtttctattggtacc 2273

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues 2688314
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_ba3.*
4: gb_in1.*
5: gb_in2.*
6: gb_in3.*
7: gb_om.*
8: gb_ov.*
9: gb_pat1.*
10: gb_pat2.*
11: gb_ph.*
12: gb_pi1.*
13: gb_pi2.*
14: gb_pi3.*
15: gb_pi4.*
16: gb_ba1.*
17: gb_ba2.*
18: em_fun.*
19: em_higo_hum.*
20: em_higo_inv.*
21: em_higo_rod.*
22: em_hig_hum1.*
23: em_hig_hum2.*
24: em_hig_hum3.*
25: em_hig_hum4.*
26: em_hig_hum5.*
27: em_hig_hum6.*
28: em_hig_hum7.*
29: em_hig_hum8.*
30: em_hig_inv1.*
31: em_hig_inv2.*
32: em_hig_other.*
33: em_hig_rod.*
34: em_hum1.*
35: em_hum2.*
36: em_hum3.*
37: em_hum4.*
38: em_hum5.*
39: em_hum6.*
40: em_hum7.*
41: em_in.*
42: em_om.*
43: em_or.*

44: em_ov.*
45: em_pat.*
46: em_ph.*
47: em_pi.*
48: em_ro.*
49: em_sts.*
50: em_sy.*
51: em_un.*
52: em_vi.*
53: gb_sts1.*
54: gb_sts2.*
55: gb_sts3.*
56: gb_sy.*
57: gb_un.*
58: gb_vil.*
59: gb_v12.*
60: gb_v12.*
61: gb_v12.*
62: gb_v12.*
63: gb_v12.*
64: gb_v12.*
65: gb_v12.*
66: gb_v12.*
67: gb_v12.*
68: gb_v12.*
69: gb_v12.*
70: gb_v12.*
71: gb_v12.*
72: gb_v12.*
73: gb_v12.*
74: gb_v12.*
75: gb_v12.*
76: gb_v12.*
77: gb_v12.*
78: gb_v12.*
79: gb_v12.*
80: gb_v12.*
81: gb_v12.*
82: gb_v12.*
83: gb_v12.*
84: gb_v12.*
85: gb_v12.*
86: gb_v12.*
87: gb_v12.*
88: gb_v12.*
89: gb_v12.*
90: gb_v12.*
91: gb_v12.*
92: gb_v12.*
93: gb_v12.*
94: gb_v12.*
95: gb_v12.*
96: gb_v12.*
97: gb_v12.*
98: em_ba3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2273	100.0	12845	9	AX002124
2	510	22.4	510	94	AF061804
3	430	18.9	430	94	AF153058
C 4	333.4	14.7	214348	66	AC021220
C 5	321	14.1	135552	65	AC019251
C 6	254.4	11.2	126330	63	AC013745
C 7	56.6	2.5	230352	64	AC016982
8	56	2.5	23108	94	AF088189

Sequence
AF061804 Mus muscu
AF153058 Mus muscu
AC021220 Homo sapi
AC019251 Homo sapi
AC013745 Homo sapi
AC016982 Mus muscu
AF088189 Mus muscu

9	56	2.5	68121	94	AP001917	Mus muscu
10	56	2.5	108333	94	AP001293	Mus muscu
11	56	2.5	213061	67	AC023248	Mus muscu
12	56	2.5	281000	94	MMU276505	Mus muscu
13	55.8	2.5	166513	75	AC073787	Mus muscu
14	55	2.4	157996	94	AC019026	Mus muscu
15	55	2.4	196858	77	AC084409	Mus muscu
16	54.4	2.4	7218	91	AP000293	Mus muscu
17	54.4	2.4	100000	91	AP000043	Mus muscu
18	54.4	2.4	100000	91	AP000111	Mus muscu
19	54.4	2.4	100000	91	AP000187	Mus muscu
20	54.4	2.4	223431	75	AC074310	Mus muscu
21	54.4	2.4	340000	91	AP001716	Mus muscu
22	54.2	2.4	207358	80	AL359736	Mus muscu
23	54.2	2.4	207430	78	AL136963	Mus muscu
24	53.4	2.3	215126	76	AC083916	Mus muscu
25	53.2	2.3	199242	81	AL589652	Mus muscu
26	53.2	2.3	204755	82	AL590390	Mus muscu
27	53	2.3	145347	75	AC073821	Mus muscu
28	53	2.3	163084	82	AL590410	Mus muscu
29	53	2.3	195556	82	AL590140	Mus muscu
30	53	2.3	209173	81	AL512597	Mus muscu
31	53	2.3	216021	97	HUAC004787	Mus muscu
32	53	2.3	234257	76	AC079574	Mus muscu
33	53	2.3	241012	76	AC079518	Mus muscu
34	53	2.3	340969	81	AL512630	Mus muscu
35	52.8	2.3	185694	76	AC079505	Mus muscu
36	52.8	2.3	213674	76	AC079528	Mus muscu
37	52.8	2.3	231669	73	AC068067	Mus muscu
38	52.4	2.3	9479	68	AC023897	Mus muscu
39	52.4	2.3	220170	74	AC073511	Mus muscu
40	52.2	2.3	179992	78	AC091251	Mus muscu
41	52.2	2.3	202083	68	AC023833	Mus muscu
42	51.8	2.3	107257	94	AF289665	Mus muscu
43	51.8	2.3	200729	67	AC022781	Mus muscu
44	51.8	2.3	200849	78	AC091250	Mus muscu
45	51.6	2.3	190430	77	AC087873	Mus muscu

ALIGNMENTS

RESULT	1					
LOCUS	AX002124	12845 bp	DNA	PAT	10-MAR-2000	
DEFINITION	Sequence 1 from Patent WO9855638.					
ACCESSION	AX002124					
VERSION	AX002124.1	GI:7241839				
KEYWORDS	unidentified.					
SOURCE	unidentified.					
ORGANISM	unclassified.					
REFERENCE	1 (bases 1 to 12845)					
AUTHORS	Breier,G. and Risau,W.					
TITLE	REGULATORY SEQUENCES CAPABLE OF CONFERRING EXPRESSION OF A HETEROLOGOUS DNA SEQUENCE IN ENDOTHELIAL CELLS IN VIVO AND USES THEREOF					
JOURNAL	Patent: WO 9855638-A 1 10-DEC-1998;					
FEATURES	MAX PLANCK GESELLSCHAFT (DE); BREIER GEORG (DE)					
source	Location/Qualifiers					
	1..12845					
	/organism="unidentified"					
	/db_xref="taxon:32644"					
BASE COUNT	3321 a 2893 c 2863 g 3723 t	45	others			
ORIGIN						
Query Match	100.0%;	Score 2273;	DB 9;	Length 12845;		
Best Local Similarity	100.0%;	Pred. No. 0;				
Matches 2273;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
QY	1	tctcgagacagcgctgtgttttccctctggtcttgcagcggaagataacagtg	60			

Db	8336	TCTCGAGACAGAGCCGCTGTTGTTTTTCTCTGGTCTTTTGGCGGGAAGATAACAGTG	8395
QY	61	cacaaataaataaattgttgcgatttgacataaaaggcttttattgtatagta	120
Db	8396	CACAATAATTANTGTTGGTATCGGATTTGAACATAAAGGGCTTTTATTGTATAGTA	8455
QY	121	gcatagtacctcttgacagcagaatgagctgctctaaagaacacccccaaacttgcga	180
Db	8456	GCATATGTACCTCTTGCAGTCAGATGAGTCTCTAAAGAACAACCCAAACTTGCCGA	8515
QY	181	tgaataagaatgaggttttaataaaaggcgatgagcatttagcactgaigtaaatctc	240
Db	8516	TGAAAATGAATGAGGTTTAAATAAGCGGATGGATGAGCATTTAGTCACTGATGATAATCTC	8575
QY	241	cagttatgtataacctcattgactgatttgattgagacatatttggtatggggcattc	300
Db	8576	CAGTTATTGATAACCTCATTGACTGGATTGTGTCAGACATGTTGGTATGGGGCATC	8635
QY	301	ctttaaagatgagcattagcacaagctgcctgcactctaaagaatactatgctgtatgta	360
Db	8636	CTTTAAAGATGAGCATAGCAACGCTGCTGCACCTCTAAGAGAATCTATGGCTGTATGTTA	8695
QY	361	ttacagagacattgagaagctcttagtgcctgctgctgctgctgctgctgctgctgctgct	420
Db	8696	TTACAGAGACAGTTGAGAAGCTCTTTAGTGGCTCTGCGTGTAGATCAGCGGTAGAGCGCT	8755
QY	421	gaggtctgcgctgcgcttctgcgactgaagaataaaggccatttactgtggtggtgcag	480
Db	8756	GAGGCTCTGCGCTCGCTTCTTGGCACTGAAGATAAAGGCCATTTACTGTGGTGGTCAG	8815
QY	481	tgggcgcgcttgcgagcttactactactactactactactactactactactactactactact	540
Db	8816	TGGGCGCAGTTTGTGACGAGTACTACTACTACTACTACTACTACTACTACTACTACTACTA	8875
QY	541	gttcacatgagatgagcgtaccagtgattgttgcaggttaattggttcagtcagtcagttta	600
Db	8876	GTTTCATCAGATCAGCGGTATCCAGTGATGTTTGCAGGTTAATGTTCTCAGTCTATGTTT	8935
QY	601	gaatcactatcaaaacaaattgtttctcatttctcctgcttcttctcaaaacaaagtaaga	660
Db	8936	GAATCTACTATCAAAACAAATTTGTTTCTCTATTCTCTGCTTCTCTCAAAACAAAGTAAGA	8995
QY	661	ttccattattgaaagctgttttaagagacattttaaactgcttgcctatgttagggacagt	720
Db	8996	TTCCATTATTGAAGAGCTTGTTTAAGAGCATTTTAACTGCTTGCCTATGTATGGGACAGT	9055
QY	721	gactatttcatattgacaaatattatgcgattatgaatgaatgactacccagttctat	780
Db	9056	GACTTATTTCATATTGCACAAATATTATGCGGATTAATTGNATATGACTACCCAGTTCTAT	9115
QY	781	agctgctcagggcagacacagacatctgtgacatccactttaaagtcatttaaaagccattaaa	840
Db	9116	AGCTGCTCAGGGCAGACCAAGAGCATCTGTGATCCAGTCACTTTAAATGCCATTTAAAA	9175
QY	841	tgcataattgtgtgcttaggaataaaacacactgtaaaagtttagaatacagggcccaaaa	900
Db	9176	TGCATAAATTGTGTGCTTAGGAATAAACACACTGTAAAGTTTGAATACAGCGGCCAAACA	9235
QY	901	caagtcttaacaaagcgaactagctctgagattcatttaattgatttatttaccatg	960
Db	9236	CAAGTCTTTAACAAATGCCAACTAGCTTCTGAGATTTCATTAATGATCAATTAATTAACCAATG	9295
QY	961	ttttaaataatgctcatttaattactaaatctatagttgtaacagcaacacatgacatct	1020
Db	9296	TTTTAAAAATATGTCATTAATTACTAAATCTATAGTTGTAAACAGCAACACATGTACATCT	9355
QY	1021	tattaagttgggtatattcaggtggcattagctgtagacattgacacatgtgtgtgtg	1080
Db	9356	TATTAAAGTTGGGTATATTTCAGGGTGGCATAGCTGTAGACTATTGGACATCTGTGTGGTG	9415
QY	1081	agccagtgagagactgcctccctggctgttctcagaaggccacagtgctcagcgattggct	1140
Db	9416	AGCCAGTGAGAACTGCCTCCTGGCTGCTTCTCAGAGGGCCACAGTGTACAGGCATGGCT	9475

* 6525 9844: contig of 3320 bp in length
 * 9845 9944: gap of unknown length
 * 9945 14101: contig of 4157 bp in length
 * 14102 14201: gap of unknown length
 * 14202 21279: contig of 7078 bp in length
 * 21280 21379: gap of unknown length
 * 21380 31346: contig of 9967 bp in length
 * 31347 31446: gap of unknown length
 * 31447 40526: contig of 9080 bp in length
 * 40527 40627: gap of unknown length
 * 40628 52863: contig of 12236 bp in length
 * 52864 60909: gap of unknown length
 * 60910 66191: gap of unknown length
 * 66192 82659: contig of 16469 bp in length
 * 82660 82759: gap of unknown length
 * 82760 104797: contig of 22038 bp in length
 * 104798 104897: gap of unknown length
 * 104898 129319: contig of 13128 bp in length
 * 129320 129420: gap of unknown length
 * 129421 176511: contig of 47092 bp in length
 * 176512 176612: gap of unknown length
 * 176613 177667: contig of 1055 bp in length
 * 177668 177767: gap of unknown length
 * 177768 178998: contig of 1231 bp in length
 * 178999 179097: gap of unknown length
 * 179098 180242: contig of 1145 bp in length
 * 180243 180342: gap of unknown length
 * 180343 181442: contig of 1100 bp in length
 * 181443 181542: gap of unknown length
 * 181543 182642: contig of 1100 bp in length
 * 182643 182742: gap of unknown length
 * 182743 184165: contig of 1423 bp in length
 * 184166 184266: gap of unknown length
 * 184267 185122: contig of 1247 bp in length
 * 185123 185613: gap of unknown length
 * 185614 187121: contig of 1509 bp in length
 * 187122 187222: gap of unknown length
 * 187223 188524: contig of 1303 bp in length
 * 188525 188624: gap of unknown length
 * 188625 189782: contig of 1158 bp in length
 * 189783 189882: gap of unknown length
 * 189883 191312: contig of 1430 bp in length
 * 191313 191412: gap of unknown length
 * 191413 192837: contig of 1425 bp in length
 * 192838 192937: gap of unknown length
 * 192939 194117: contig of 1180 bp in length
 * 194118 194217: gap of unknown length
 * 194218 195948: contig of 1731 bp in length
 * 195949 196048: gap of unknown length
 * 196049 197744: contig of 1696 bp in length
 * 197745 197844: gap of unknown length
 * 197845 199489: contig of 1645 bp in length
 * 199490 199589: gap of unknown length
 * 199590 201304: contig of 1715 bp in length
 * 201305 201404: gap of unknown length
 * 201405 203316: contig of 1912 bp in length
 * 203317 203416: gap of unknown length
 * 203417 204490: contig of 1074 bp in length
 * 204491 204590: gap of unknown length
 * 204591 206676: contig of 2086 bp in length
 * 206677 206776: gap of unknown length
 * 206777 208803: contig of 2027 bp in length
 * 208804 208903: gap of unknown length
 * 208904 210597: contig of 1694 bp in length
 * 210598 210697: gap of unknown length
 * 210698 212097: contig of 1400 bp in length
 * 212098 212197: gap of unknown length
 * 212198 214348: contig of 2151 bp in length.

FEATURES
 source

/chromosome="4"
 /clone="Rp11-530117"
 1..1584
 /note="assembly_name:Contig100"
 1685..4194
 /note="assembly_name:Contig101"
 4295..6424
 /note="assembly_name:Contig102"
 6525..9844
 /note="assembly_name:Contig103"
 9945..14101
 /note="assembly_name:Contig104"
 14202..21279
 /note="assembly_name:Contig105"
 21380..31346
 /note="assembly_name:Contig106"
 31447..40526
 /note="assembly_name:Contig107"
 clone_end:77
 vector_side:left"
 40627..52862
 /note="assembly_name:Contig108"
 52963..66090
 /note="assembly_name:Contig109"
 66191..82659
 /note="assembly_name:Contig110"
 82760..104797
 /note="assembly_name:Contig111"
 104898..129319
 /note="assembly_name:Contig112"
 clone_end:SP6
 vector_side:right"
 129420..176511
 /note="assembly_name:Contig113"
 176612..177666
 /note="assembly_name:Contig133"
 177767..178997
 /note="assembly_name:Contig37"
 179098..180242
 /note="assembly_name:Contig55"
 180343..181442
 /note="assembly_name:Contig59"
 181543..182642
 /note="assembly_name:Contig66"
 182743..184165
 /note="assembly_name:Contig69"
 184266..185512
 /note="assembly_name:Contig75"
 185613..187121
 /note="assembly_name:Contig77"
 187222..188524
 /note="assembly_name:Contig78"
 188625..189782
 /note="assembly_name:Contig80"
 189883..191312
 /note="assembly_name:Contig85"
 191413..192837
 /note="assembly_name:Contig86"
 192938..194117
 /note="assembly_name:Contig87"
 194218..195948
 /note="assembly_name:Contig88"
 196049..197744
 /note="assembly_name:Contig89"
 197845..199489
 /note="assembly_name:Contig90"
 199590..201304
 /note="assembly_name:Contig91"
 201405..203316
 /note="assembly_name:Contig93"
 203417..204490

Query Match

14.7%; Score 333.4; DB 66; Length 214348;

RESULT 6
AC013745/c

LOCUS

DEFINITION

AC013745 126330 bp DNA HTG 13-JUL-2000

SAMPLING

Homo sapiens chromosome 4 clone RP11-92P18 map 4, LOW-PASS SEQUENCE

ACCESSION

AC013745

VERSION

AC013745.3

KEYWORDS

HTG; HTGS_PHASE0.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 126330)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,

Baldwin, J., Castle, A., Collins, S., Collins, S., Collins, S., Collins, S.,

Cooke, P., DeArrellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,

Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,

Galaan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,

Lehoczky, J., Liu, C., Locke, K., Macdonald, P., Marquis, N.,

McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,

Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,

Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,

Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,

Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 13, 2000 this sequence version replaced gi:6910837.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L2025

Center clone name: 92_P_18

* NOTE: This record contains 147 individual

* sequencing reads that have not been assembled into

* contigs. Runs of N are used to separate the reads

* and the order in which they appear is completely

* arbitrary. Low-pass sequence sampling is useful for

* identifying clones that may be gene-rich and allows

* overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone

* will be sequenced to completion. In the event that

* the record is updated, the accession number will

* be preserved.

* 1 722: contig of 722 bp in length

* 723 822: gap of 100 bp

* 823 1542: contig of 720 bp in length

* 1543 1642: gap of 100 bp

* 1643 2371: contig of 729 bp in length

* 2372 2471: gap of 100 bp

* 2472 3212: contig of 741 bp in length

* 3213 3312: gap of 100 bp

* 3313 4049: contig of 737 bp in length

* 4050 4149: gap of 100 bp

* 4150 4840: contig of 691 bp in length

* 4841 4940: gap of 100 bp

* 4941 5676: contig of 736 bp in length

* 5677 5776: gap of 100 bp

* 5777 6521: contig of 745 bp in length

* 6522 6621: gap of 100 bp

* 6622 7344: contig of 723 bp in length

* 7345 7444: gap of 100 bp

* 7445 8161: contig of 717 bp in length

* 8162 8261: gap of 100 bp

* 8262 9014: contig of 753 bp in length

* 9015 9114: gap of 100 bp

* 9115 9827: contig of 713 bp in length

* 9828 9927: gap of 100 bp

* 9928 10650: contig of 723 bp in length

* 10651 10750: gap of 100 bp

* 10751 11479: contig of 729 bp in length

* 11480 11579: gap of 100 bp

* 11580 12318: contig of 739 bp in length

* 12319 12418: gap of 100 bp

* 12419 13137: contig of 719 bp in length

* 13138 13237: gap of 100 bp

* 13238 13959: contig of 722 bp in length

* 13960 14059: gap of 100 bp

* 14060 14785: contig of 726 bp in length

* 14786 14885: gap of 100 bp

* 14886 15580: contig of 695 bp in length

* 15581 15680: gap of 100 bp

* 15681 16420: contig of 740 bp in length

* 16421 16520: gap of 100 bp

* 16521 17233: contig of 713 bp in length

* 17234 17333: gap of 100 bp

* 17334 18057: contig of 724 bp in length

* 18058 18157: gap of 100 bp

* 18158 18874: contig of 717 bp in length

* 18875 18974: gap of 100 bp

* 18975 19716: contig of 742 bp in length

* 19717 19816: gap of 100 bp

* 19817 20557: contig of 741 bp in length

* 20558 20657: gap of 100 bp

* 20658 21379: contig of 722 bp in length

* 21380 21479: gap of 100 bp

* 21480 22193: contig of 714 bp in length

* 22194 22293: gap of 100 bp

* 22294 23013: contig of 720 bp in length

* 23014 23113: gap of 100 bp

* 23114 23851: contig of 738 bp in length

* 23852 23951: gap of 100 bp

* 23952 24681: contig of 730 bp in length

* 24682 24781: gap of 100 bp

* 24782 25505: contig of 724 bp in length

* 25506 25605: gap of 100 bp

* 25606 26335: contig of 730 bp in length

* 26336 26435: gap of 100 bp

* 26436 27158: contig of 723 bp in length

* 27159 27258: gap of 100 bp

* 27259 27960: contig of 702 bp in length

* 27961 28060: gap of 100 bp

* 28061 28775: contig of 715 bp in length

* 28776 28875: gap of 100 bp

* 28876 29586: contig of 711 bp in length

* 29587 29686: gap of 100 bp

* 29687 30394: contig of 708 bp in length

* 30395 30494: gap of 100 bp

* 30495 31246: contig of 752 bp in length

* 31247 31346: gap of 100 bp

* 31347 32090: contig of 744 bp in length

* 32091 32190: gap of 100 bp

* 32191 32894: contig of 704 bp in length

* 32895 32994: gap of 100 bp

* 32995 33741: contig of 747 bp in length

* 33742 33841: gap of 100 bp

* 33842 34602: contig of 761 bp in length

* 34603 34702: gap of 100 bp

* 34703 35440: contig of 738 bp in length

* 35441 35540: gap of 100 bp

* 35541 36261: contig of 721 bp in length

* 36262 36361: gap of 100 bp

36362 37081: contig of 720 bp in length
37082 37181: gap of 100 bp
37182 37898: contig of 717 bp in length
37899 37998: gap of 100 bp
37999 38730: contig of 732 bp in length
38731 38930: gap of 100 bp
38931 39526: contig of 696 bp in length
39527 39626: gap of 100 bp
39627 40316: contig of 690 bp in length
40317 40416: gap of 100 bp
40417 41136: contig of 720 bp in length
41137 41236: gap of 100 bp
41237 41985: contig of 749 bp in length
41986 42085: gap of 100 bp
42086 42815: contig of 730 bp in length
42816 42915: gap of 100 bp
42916 43632: contig of 717 bp in length
43633 43732: gap of 100 bp
43733 44445: contig of 713 bp in length
44446 44545: gap of 100 bp
44546 45287: contig of 742 bp in length
45288 45387: gap of 100 bp
45388 46129: contig of 742 bp in length
46130 46229: gap of 100 bp
46230 46961: contig of 732 bp in length
46962 47061: gap of 100 bp
47062 47765: contig of 704 bp in length
47766 47865: gap of 100 bp
47866 48582: contig of 717 bp in length
48583 48682: gap of 100 bp
48683 49429: contig of 747 bp in length
49430 49529: gap of 100 bp
49530 50236: contig of 707 bp in length
50237 50336: gap of 100 bp
50337 51062: contig of 726 bp in length
51063 51162: gap of 100 bp
51163 51884: contig of 722 bp in length
51885 51984: gap of 100 bp
51985 52702: contig of 718 bp in length
52703 52802: gap of 100 bp
52803 53516: contig of 714 bp in length
53517 53616: gap of 100 bp
53617 54329: contig of 713 bp in length
54330 54429: gap of 100 bp
54430 55140: contig of 711 bp in length
55141 55240: gap of 100 bp
55241 55930: contig of 690 bp in length
55931 56030: gap of 100 bp
56031 56724: contig of 694 bp in length
56725 56824: gap of 100 bp
56825 57628: contig of 804 bp in length
57629 57728: gap of 100 bp
57729 58523: contig of 795 bp in length
58524 58623: gap of 100 bp
58624 59387: contig of 764 bp in length
59388 59487: gap of 100 bp

Query Match 11.2%; Score 254.4; DB 63; Length 126330;
Best Local Similarity 66.4%; Pred. No. 3.2e-53;
Matches 314; Conservative 0; Mismatches 144; Indels 15; Gaps 2;
QY 1815 aggaagtctggtacacagcatgataaaagacaatggagcggtggtcagtggtccgcgtc 1874
|||
Db 88902 ANNAACAAAGTTATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 88843
QY 1875 ccttcagggtgtagagacagctgtagagatgtctccaggaggttttcaatca 1934
|||||
Db 88842 NNAATCA 88783
QY 1935 gcaatttagctagctgtgcatctctattacagaataatcagtggtggcc----- 1987
|||||
Db 88782 GCAATTAGTCAAAATGTGACATCCCTATGTCTATAGAAAGATGTCACTGGTCCCTTTCC 88723

QY 1988 -----tgagatcatcagatggaggttcacatcggtgttcacgtcccggtatccctttgta 2040
|||||
Db 88722 AAGGAGTGTGATCATCAGATGAAGTTTCATTTGGTTTCATGTCCTCGTATCCTTTGTA 88563
QY 2041 agaccttgaagtgtggcaacgaggaacacaggaactcacccctcgtggtgcgtggaattgcag 2100
|||||
Db 88662 AGACCTTGAAGTTGGCAATGCAGGAAACAGAGAACTCACCTAGCTCCATGAATTGCAG 88603
QY 2101 agctgttgt 2160
|||||
Db 88602 AACTGT 88543
QY 2161 ctttaactgggactgggggcaagtcac-tccacactttatacaatgaattctgaagagg 2219
|||||
Db 88542 CTTTACTGAGATGGTGAAGTAAATCCACGTTTATACATGAATTCCTGAGAGAGG 88483
QY 2220 ccttttaactggagtgatgt 2272
|||||
Db 88482 CTTTAAAGTATAGATGATGCAATGTTTATGGAAGGTGTTTCTCTATTAGTGC 88430
AC016982 230352 bp DNA HTG 26-FEB-2001
Mus musculus clone RP23-416H2, WORKING DRAFT SEQUENCE, 11 unordered
pieces.
AC016982
AC016982.4 GI:13123354
HTG; HTGS_PHASE1; HTGS_DRAFT.
house mouse.
Mus musculus
ORGANISM
REFERENCE
1 (bases 1 to 230352)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE
Mus musculus, clone RP23-416H2
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 230352)
AUTHORS
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,
Bohuslavsky, L., Bouckgalter, B., Brown, A., Castle, A., Colangelo, M.,
Collins, S., Collymore, A., Cooke, P., DeArellano, K., Dewar, K.,
Domino, M., Doyle, M., Fenestor, J., Ferreira, P., FitzHugh, W.,
Forrest, C., Gage, D., Galastor, J., Gardyna, S., Grant, G., Hagos, B.,
Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L.,
Karatas, A., Klein, J., Landers, T., Lehoczy, J., Lieu, C., Lock, K.,
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
Meldrim, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T.,
O'Donnell, P., Peterson, K., Pierre, N., Pollara, V., Riley, R.,
Rothman, D., Roy, A., Santos, R., Severy, P., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
Zimmer, A. and Zody, M.
Direct Submission
TITLE
JOURNAL
COMMENT
Submitted (09-DEC-1999), Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 25, 2001 this sequence version replaced gi:8072589.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5038
Center clone name: 416_H_2
----- Summary Statistics
Sequencing vector: M13; M77815; 39% of reads
Sequencing vector: pLasmid; n/a; 61% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731


```
QY 1355 acttgccggtccattgtctcactcactcagaatgctgggtccaggtgtgtgaccacac 1414
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3086 ACTCTCTTCTGCTAGTCTAGGCTCCAGGACTGGGGTTACAGGTGTGAGGCGACAA 3145
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1415 taagtactgcggttttaagactaagagctgaagatcctgatgctcctt 1462
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3146 AGCCAGTTTCTAGTCTTCTAGTCTATATCAAAACATTAATGCATT 3193
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
AP001917 68121 bp DNA ROD 21-OCT-2000
LOCUS Mus musculus genomic DNA , chromosome 7, clone:B131T.
DEFINITION AP001917
ACCESSION AP001917
VERSION AP001917.2 GI:10945234
KEYWORDS
SOURCE Mus musculus DNA, clone:B131T.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 68121)
AUTHORS Hattori,M., Watanabe,H., Toyoda,A. and Sakaki,Y.
TITLE house mouse 68,121 genomic DNA
JOURNAL Published Only in DataBase (2000) In press
REFERENCE 2 (bases 1 to 68121)
AUTHORS Hattori,M., Watanabe,H., Toyoda,A. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2000) to the DDBJ/EMBL/GenBank databases.
Masahira Hattori, The Institute of Physical and Chemical Research
(RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,
Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Oct 21, 2000 this sequence version replaced gi:8953895.
This work was done in collaboration with Dr. Hitomi Yatsuki, Dr.
Masayo Nishimura and Dr. Tsunehiro Mukai. (Saga Medical School,
Department of Biochemistry, Saga, Saga, 849-8501, Japan).
FEATURES
source
1. .68121
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="7"
/clone="B131T"
BASE COUNT 16012 a 17409 c 17842 g 16858 t
ORIGIN

Query Match 2.5%; Score 56; DB 94; Length 68121;
Best Local Similarity 54.3%; Pred. No. 0.0026;
Matches 113; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 1327 ctcaactgtgcccaggctagtctcaaaactgcggtccattgtctcactcactcagaat 1386
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 54164 CTCACATGTAGCCAGGCGAGTCTCTAGCTTAAGGTCCTCTTGCCTCAGCCTCCTAAAT 54223
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1387 gctgggtcttcagggtgtgacacactaggtagctgcgcttttaagctaaagctgga 1446
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 54223 GCTGGGATTCAGGTGTATATCCCGGCGAGTCTCTGCAACTGTTCAAACTGGAATCTTTA 54283
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1447 agatcctgatgctccttaccatggtgggcatgttacaggttagttgactgaaactagtt 1506
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 54283 ACCACAATGAGATGTTTAAGAGGCCAGCAATATGAGAGTGTAGAAACAGAACATTCGAA 54342
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1507 atctcgtgtgtaagtacctgcagtggt 1534
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 54343 ATGTGGCTCAATAAAGCACCTAACCGGT 54370
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
AC023248/c 213061 bp DNA HTG 10-JAN-2001
LOCUS Mus musculus clone RP23-124B2, WORKING DRAFT SEQUENCE, 8 unordered
DEFINITION pieces.
ACCESSION AC023248
VERSION AC023248.3 GI:12061504
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 213061)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP23-124B2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 213061)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

QY 1355 acttgccggtccattgtctcactcactcagaatgctgggtccaggtgtgtgaccacac 1414
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3086 ACTCTCTTCTGCTAGTCTAGGCTCCAGGACTGGGGTTACAGGTGTGAGGCGACAA 3145
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1415 taagtactgcggttttaagactaagagctgaagatcctgatgctcctt 1462
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3146 AGCCAGTTTCTAGTCTTCTAGTCTATATCAAAACATTAATGCATT 3193
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
AP001917 68121 bp DNA ROD 21-OCT-2000
LOCUS Mus musculus genomic DNA , chromosome 7, clone:B131T.
DEFINITION AP001917
ACCESSION AP001917
VERSION AP001917.2 GI:10945234
KEYWORDS
SOURCE Mus musculus DNA, clone:B131T.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 68121)
AUTHORS Hattori,M., Watanabe,H., Toyoda,A. and Sakaki,Y.
TITLE house mouse 68,121 genomic DNA
JOURNAL Published Only in DataBase (2000) In press
REFERENCE 2 (bases 1 to 68121)
AUTHORS Hattori,M., Watanabe,H., Toyoda,A. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2000) to the DDBJ/EMBL/GenBank databases.
Masahira Hattori, The Institute of Physical and Chemical Research
(RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,
Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Oct 21, 2000 this sequence version replaced gi:8953895.
This work was done in collaboration with Dr. Hitomi Yatsuki, Dr.
Masayo Nishimura and Dr. Tsunehiro Mukai. (Saga Medical School,
Department of Biochemistry, Saga, Saga, 849-8501, Japan).
FEATURES
source
1. .68121
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="7"
/clone="B131T"
BASE COUNT 16012 a 17409 c 17842 g 16858 t
ORIGIN

Query Match 2.5%; Score 56; DB 94; Length 68121;
Best Local Similarity 54.3%; Pred. No. 0.0026;
Matches 113; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 1327 ctcaactgtgcccaggctagtctcaaaactgcggtccattgtctcactcactcagaat 1386
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 54164 CTCACATGTAGCCAGGCGAGTCTCTAGCTTAAGGTCCTCTTGCCTCAGCCTCCTAAAT 54223
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1387 gctgggtcttcagggtgtgacacactaggtagctgcgcttttaagctaaagctgga 1446
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 54224 GCTGGGATTCAGGTGTATATCCCGGCGAGTCTCTGCAACTGTTCAAACTGGAATCTTTA 54283
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1447 agatcctgatgctccttaccatggtgggcatgttacaggttagttgactgaaactagtt 1506
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 54284 ACCACAATGAGATGTTTAAGAGGCCAGCAATATGAGAGTGTAGAAACAGAACATTCGAA 54343
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1507 atctcgtgtgtaagtacctgcagtggt 1534
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 54344 ATGTGGCTCAATAAAGCACCTAACCGGT 54371
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
AP001293 108333 bp DNA ROD 05-DEC-2000
LOCUS Mus musculus genomic DNA, chromosome 7, clone:B131, complete
DEFINITION
```



```

Berlin, GERMANY
Revised by author 22-MAY-2000
3 (bases 1 to 281000)
Engemann, S.
Direct Submission
Submitted (22-NOV-2000) Engemann S., Dep. Trautner,
Max-Planck-Institut fuer Molekulare Genetik, Ihnestrasse 73, 14195
Berlin, GERMANY
4 (bases 1 to 281000)
Engemann, S.
Direct Submission
Submitted (24-JAN-2001) Engemann S., Dep. Trautner,
Max-Planck-Institut fuer Molekulare Genetik, Ihnestrasse 73, 14195
Berlin, GERMANY
On Jan 27, 2001 this sequence version replaced gi:11191799.
related accession numbers AJ300452-AJ300460
has a 1000 bp overlap with AJ271885.

FEATURES
    source
        Location/Qualifiers
            1..281000
                /organism="Mus musculus domesticus"
                /strain="129 SV"
                /sub_species="domesticus"
                /db_xref="taxon:10092"
                /chromosomes="7"
                /cell_line="CJ7"
            join(<1431..1455,3536..3622,7265..7345,7478..7578,
7906..8110,9688..9772,11978..12149,12416..12608,
13783..13967,14347..14428,15828..15926,16308,16404,
17790..17888,21514..21610,21996..22086,22181..22317,
22743..22928,23228..23354,24142..24268,25136..25238,
27169..28364)
                /gene="Obph1"
            1431..28364
                /gene="Obph1"
            1432..1455
                /gene="Obph1"
            /number=1
            join(1432..1455,3536..3622,7265..7345,7478..7578,
7906..8110,9688..9772,11978..12149,12416..12608,
13783..13967,14347..14428,15828..15926,16308,16404,
17790..17888,21514..21610,21996..22086,22181..22317,
22743..22928,23228..23354,24142..24268,25136..25238,
27169..27307)
                /gene="Obph1"
            /codon_start=1
            /product="oxysterol binding protein"
            /protein_id="CAC27351.1"
            /db_xref="GI:12583596"

    mRNA
        130688,131306..132456,
140377..140517,141097..141184,142468..142590,
145511..145624,146438..146585,146958..147085,
147169..147274,147610..147707,147985..148174,
148724..148792,152041..152121,154073..154154,
155189..155255,157406..157466,157904..157988,
159108..159408)
                /gene="Cars"
            30593..130688
                /gene="Cars"
            /number=5
            join(<31277..31342,35572..35662,37049..37176,46179..46291,
47068..47141,48531..48627)
                /gene="Tnfrh1"
            /join(31277..31342,35572..35662,37049..37176,46179..46291,
47068..47141,48531..48627)
                /gene="Tnfrh1"
            35584..35662
                /gene="Tnfrh1"
            /number=1
            join(35584..35662,37049..37176,46179..46291,47068..47141,
48531..48580)
                /gene="Tnfrh1"

    exon
        9688..9772
            /gene="Obph1"
            /number=5
        11978..12149
            /gene="Obph1"
            /number=6
        12416..12608
            /gene="Obph1"
            /number=7
        13783..13967
            /gene="Obph1"
            /number=8
        14347..14428
            /gene="Obph1"
            /number=9
        15828..15926
            /gene="Obph1"
            /number=10
        16308..16404
            /gene="Obph1"
            /number=11
        17790..17888
            /gene="Obph1"
            /number=12
        21514..21610
            /gene="Obph1"
            /number=13
        21996..22086
            /gene="Obph1"
            /number=14
        22181..22317
            /gene="Obph1"
            /number=15
        22743..22928
            /gene="Obph1"
            /number=16
        23228..23354
            /gene="Obph1"
            /number=17
        24142..24268
            /gene="Obph1"
            /number=18
        25136..25238
            /gene="Obph1"
            /number=19
        27169..27307
            /gene="Obph1"
            /number=20
        /number=21
        join(30593..130688,131306..132456,
140377..140517,141097..141184,142468..142590,
145511..145624,146438..146585,146958..147085,
147169..147274,147610..147707,147985..148174,
148724..148792,152041..152121,154073..154154,
155189..155255,157406..157466,157904..157988,
159108..159408)
            /gene="Cars"
        30593..130688
            /gene="Cars"
            /number=5
        join(<31277..31342,35572..35662,37049..37176,46179..46291,
47068..47141,48531..48627)
            /gene="Tnfrh1"
        /join(31277..31342,35572..35662,37049..37176,46179..46291,
47068..47141,48531..48627)
            /gene="Tnfrh1"
        35584..35662
            /gene="Tnfrh1"
            /number=1
        join(35584..35662,37049..37176,46179..46291,47068..47141,
48531..48580)
            /gene="Tnfrh1"

    CDS
        9688..9772
            /gene="Obph1"
            /number=5
        11978..12149
            /gene="Obph1"
            /number=6
        12416..12608
            /gene="Obph1"
            /number=7
        13783..13967
            /gene="Obph1"
            /number=8
        14347..14428
            /gene="Obph1"
            /number=9
        15828..15926
            /gene="Obph1"
            /number=10
        16308..16404
            /gene="Obph1"
            /number=11
        17790..17888
            /gene="Obph1"
            /number=12
        21514..21610
            /gene="Obph1"
            /number=13
        21996..22086
            /gene="Obph1"
            /number=14
        22181..22317
            /gene="Obph1"
            /number=15
        22743..22928
            /gene="Obph1"
            /number=16
        23228..23354
            /gene="Obph1"
            /number=17
        24142..24268
            /gene="Obph1"
            /number=18
        25136..25238
            /gene="Obph1"
            /number=19
        27169..27307
            /gene="Obph1"
            /number=20
        /number=21
        join(30593..130688,131306..132456,
140377..140517,141097..141184,142468..142590,
145511..145624,146438..146585,146958..147085,
147169..147274,147610..147707,147985..148174,
148724..148792,152041..152121,154073..154154,
155189..155255,157406..157466,157904..157988,
159108..159408)
            /gene="Cars"
        30593..130688
            /gene="Cars"
            /number=5
        join(<31277..31342,35572..35662,37049..37176,46179..46291,
47068..47141,48531..48627)
            /gene="Tnfrh1"
        /join(31277..31342,35572..35662,37049..37176,46179..46291,
47068..47141,48531..48627)
            /gene="Tnfrh1"
        35584..35662
            /gene="Tnfrh1"
            /number=1
        join(35584..35662,37049..37176,46179..46291,47068..47141,
48531..48580)
            /gene="Tnfrh1"

```

```

/codon_start=1
/product="tumor necrosis factor receptor p60 homologue"
/protein_id="CAC27352.1"
/db_xref="GI:12583597"
/translation="MPESYFNGCPDGEYQSDYCCKTCPSGTFVKAPCKIPHTOGQCE
KCHPCYFTGCKDNLHDLCELCSTCDKDONMWADCSATSDRKCEQIGLYYDPKPPESC
RPTCKCPQGIPIVQECNSTANTVCCSSVSNPRNLFWLMLIVFCI"
exon      37049..37176
           /gene="Tnfrhl"
           /number=2
exon      46179..46291
           /gene="Tnfrhl"
           /number=3
exon      47068..47141
           /gene="Tnfrhl"
           /number=4
exon      48531..48580
           /gene="Tnfrhl"
           /number=5
mRNA      join(567525..67596,72244..72334,73711..73838,76251..76363,
77140..77213,77826..77868,78677..78828)
           /gene="Tnfrh2"
gene       join(67525..67596,72244..72334,73711..73838,76251..76363,
77140..77213,77826..77868,78677..78828)
           /gene="Tnfrh2"
CDS       join(72256..72334,73711..73838,76251..76363,77140..77213,
77826..77857)
Query Match      2.5%; Score 56; DB 94; Length 281000;
Best Local Similarity 54.3%; Pred. No. 0.0028;
Matches 113; Conservative 0; Mismatches 95; Indels 0; Gaps 0;
QY 1327 ctcaactgtggccaggtagctcacaactggtcggtcattctctcaactcagaatt 1386
Db 275809 CTCACATATGAGCGAGCGAGTCTAGCTTAGGCTCTTGCTCAGCTCTCTAAT 275868
QY 1387 gctggggttcagggtgtgtgcaccacactaggtagctgcgtctttaagctgaagctgga 1446
Db 275869 GCTGGGATTCAGGTATATCTCCGGCAGTCTCTGCAACTGTTCAAACTGGAAATCTTTA 275928
QY 1447 agatcctgatgtccttaccatggtggcagtttacaggttagttgactgaaactagtt 1506
Db 275929 ACCACATGAGAGTGTAAAGAGGCCAGAGATATGAGAGTGTAGAAACAGACATTCGAA 275988
QY 1507 atctcgtgtgtaatgacctgcagtggt 1534
Db 275989 ATGTGGCTCAATAAAGCACCTAACCGGT 276016

```

RESULT 13

```

AC073787
LOCUS      AC073787.2 166513 bp DNA HTG 18-JUL-2000
DEFINITION Mus musculus clone RP23-413L18, WORKING DRAFT SEQUENCE, 16 ordered
pieces.
ACCESSION AC073787
VERSION    AC073787.2 GI:9256796
KEYWORDS   HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 166513)
            DOE Joint Genome Institute.
            Sequencing of Mouse
            Unpublished
REFERENCE  2 (bases 1 to 166513)
            DOE Joint Genome Institute.
            Direct Submission
            Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            On Jul 18, 2000 this sequence version replaced gi:8810404.
            -----Genome Center
            Center: Joint Genome Institute

```

```

Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1895746
Center clone name: RPCI-23_413L18
-----
Summary Statistics
Consensus quality: 157635 bases at least Q40
Consensus quality: 163578 bases at least Q30
Consensus quality: 164611 bases at least Q20
Estimated insert size: 183000; agarose-fp estimation
Estimated insert size: 165813; sum-of-contigs estimation
Quality coverage: 6.23 in Q20 bases; agarose-fp estimation
Quality coverage: 6.88 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 2214: contig of 2214 bp in length
* 2215 15749: contig of 13435 bp in length
* 2315 15849: gap of unknown length
* 15750 31759: contig of 15910 bp in length
* 15850 31759: gap of unknown length
* 31760 31859: gap of unknown length
* 31860 47298: contig of 15439 bp in length
* 47299 47398: gap of unknown length
* 47399 50191: contig of 2793 bp in length
* 50192 50291: gap of unknown length
* 50292 53149: contig of 2858 bp in length
* 53150 53250: gap of unknown length
* 53250 64766: contig of 11517 bp in length
* 64767 64867: gap of unknown length
* 64867 67071: contig of 2204 bp in length
* 67071 67171: gap of unknown length
* 67171 87473: contig of 20303 bp in length
* 87474 87573: gap of unknown length
* 87574 97882: contig of 10109 bp in length
* 97883 97882: gap of unknown length
* 97883 115608: contig of 17826 bp in length
* 115609 127399: contig of 11691 bp in length
* 127400 127499: gap of unknown length
* 127500 128951: contig of 1452 bp in length
* 128952 129051: gap of unknown length
* 129052 149900: contig of 20848 bp in length
* 149900 149999: gap of unknown length
* 150000 151700: contig of 1701 bp in length
* 151701 151800: gap of unknown length
* 151801 166513: contig of 14713 bp in length.
FEATURES
            Location/Qualifiers
            1..166513
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="RP23-413L18"
            /clone_lib="RPCI mouse BAC library 23"
BASE COUNT 39178 a 43350 c 43117 g 39368 t 1500 others
ORIGIN

```

```

Query Match      2.5%; Score 55.8; DB 75; Length 166513;
Best Local Similarity 58.0%; Pred. No. 0.0031;
Matches 120; Conservative 0; Mismatches 82; Indels 5; Gaps 1;
QY 1285 tgccctgatttaattagttattctctctctgagacagagctcactgtgtgcccagg 1344
Db 82983 TGAGTGCACATTTCTCTTTCTTTTGTTCCTCAGACAGACTCCCTATGTAACCTAGG 83042

```


Oy 1345 ctatgtctcaaacctgcggtccattgtctcactcatcagaatgctgggcttcacaggtgtg 1404
 Db 83043 CTGGCTCCAACTAGTATCTCTGCTCAGCTCCAGCTTCCAAATGCTGGGATTACAAATATG 83102
 Oy 1405 tgcaccacactagtagctgcgttttaagctaaagactgaagactcgtatgctcttta 1464
 Db 83103 TG-----CCTTAGCTGGCCCCAAGGATATAAATCTACAGTGGCAGTGGCTAATCTTTTGA 83157

Oy 1465 ccattgtgggcatgtttacaggttagtt 1491

Db 83158 AGTAGGGGTGGGTGTTGTTGTTGTT 83184

RESULT 14

AC019026

LOCUS

DEFINITION AC019026 157996 bp DNA ROD 14-JUL-2000
 Mus musculus chromosome 6 clone RP23-188E5 strain C57BL/6J,
 complete sequence.

AC019026

VERSION AC019026.12 GI:8927595

KEYWORDS HTG.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS

1 (bases 1 to 157996)
 Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
 Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A. and
 Kucherlapati,R.

TITLE

JOURNAL Mouse High Throughput Sequencing

REFERENCE

2 (bases 1 to 157996)
 Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
 Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A. and
 Kucherlapati,R.

AUTHORS

TITLE

JOURNAL Direct Submission
 Submitted (30-DEC-1999) Department of Molecular Genetics, Albert
 Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
 Bronx, NY 10461, USA

REFERENCE

3 (bases 1 to 157996)
 Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
 Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A. and
 Kucherlapati,R.

AUTHORS

TITLE

JOURNAL Direct Submission
 Submitted (06-JUL-2000) Department of Molecular Genetics, Albert
 Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
 Bronx, NY 10461, USA

REFERENCE

4 (bases 1 to 157996)
 Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
 Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A. and
 Kucherlapati,R.

AUTHORS

TITLE

JOURNAL Direct Submission
 Submitted (14-JUL-2000) Department of Molecular Genetics, Albert
 Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
 Bronx, NY 10461, USA

COMMENT

On Jul 6, 2000 this sequence version replaced gi:7767645.
 -----Genome Center Center:
 Albert Einstein College of Medicine Center
 Code: AECOM
 Web site:
 http://sequence.aecom.yu.edu/cgi-
 bin/ws.exe/mouseDB/mouseSEQ/mouseseqtable.hts
 Contact: jhan@sequence.aecom.yu.edu

CLONE LENGTH: This sequence represents the entire insert of this
 clone unless otherwise noted. If there are overlapping clones, the
 overlaps are noted in the beginning and end of the Features
 listing.

ANNOTATION OF FEATURES:
 STSs are identified using ePCR (Genome Res. 7:541-550).
 Repeats are identified using RepeatMasker (A. Smit and P. Green,
 unpublished.) for Human and Mouse sequences.

Genes and Regions of sequence similarity are identified by BLAST
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to EST
 and cDNA sequences in Unigene. Genes demonstrate at least two exons
 flanked by consensus splice sites that maintain sequence continuity
 across the splice junctions. Sequences that are not identical
 matches are annotated as similar.

SEQUENCING READ COVERAGE: Attempts are made to complete double
 stranded sequence for all regions. All sequence is completed to a
 standard of coverage with a minimum of 3 reads with no ambiguities.
 If the sequence coverage for a region does not meet this standard,
 it is indicated in the annotation as Low Coverage. Low coverage
 linkages are verified by PCR product size verification or
 verification of forward and reverse reads from clones which span
 the low coverage area.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
 standards - estimated average error rate is less than 1 per 10,000
 bases using the Consed quality parameters. Regions that do not
 meet this requirement are annotated as Low Quality.

-----Summary Statistics:

Center project name: AAP
 Sequencing vector: pUC18; L08752
 Chemistry: Dye-terminator Big Dye; 100%
 Assembly program: Phrap version 0.990319
 Contig length: 157996
 Fraction of Phrap value < 40: 0.0011076
 Error rate shown by Consed: 0.13 per 10,000 bases
 Number of N's in consensus: 1

----- Distribution of Quality < 40 Bases:

	1000	900	800	700	600	500	400	300	200	100	0
# bases	1000	900	800	700	600	500	400	300	200	100	0

Phrap Value Range

FEATURES	source	Location/Qualifiers
repeat_region	repeat_region	1. 157996 /organism="Mus musculus" /db_xref="taxon:10090" /chromosome="6" /clone="RP23-188E5" 448..743 rpt_family="(TG)n" complement(2116..2138) /rpt_family="ID3" complement(2139..2274) /rpt_family="B1_MM" complement(2275..2348) /rpt_family="ID3" complement(3344..3682) /rpt_family="L1_MM" 4251..4347 /rpt_family="CT-rich" 4348..4459 /rpt_family="(CATA)n" 4631..4754 /rpt_family="B4A" 4807..4841 /rpt_family="(CATA)n" 4863..4884

Query Match 2.4%; Score 55; DB 94; Length 157996;
Best Local Similarity 66.4%; pred. No. 0.0049;
Matches 79; Conservative 0; Mismatches 40; Indels 0;

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 30351: contig of 30351 bp in length
 * 30352 30451: gap of unknown length
 * 30452 48922: contig of 18471 bp in length
 * 48923 49022: gap of unknown length
 * 49023 63234: contig of 14212 bp in length
 * 63235 63334: gap of unknown length
 * 63335 74942: contig of 11608 bp in length
 * 74943 75042: gap of unknown length
 * 75043 88417: contig of 13375 bp in length
 * 88418 88517: gap of unknown length
 * 88519 96823: contig of 8306 bp in length
 * 96824 96923: gap of unknown length
 * 96924 105059: contig of 8146 bp in length
 * 105070 105169: gap of unknown length
 * 105170 112290: contig of 7121 bp in length
 * 112291 112390: gap of unknown length
 * 112391 119958: contig of 7568 bp in length
 * 119959 120059: gap of unknown length
 * 120059 126582: contig of 6524 bp in length
 * 126583 131525: contig of 4843 bp in length
 * 131526 131626: gap of unknown length
 * 131626 138501: contig of 6876 bp in length
 * 138502 143451: gap of unknown length
 * 143452 143552: gap of unknown length
 * 143552 147862: contig of 4311 bp in length
 * 147863 152648: gap of unknown length
 * 152649 152748: gap of unknown length
 * 152749 157056: contig of 4308 bp in length
 * 157057 157157: gap of unknown length
 * 157157 161317: contig of 4161 bp in length
 * 161318 161418: gap of unknown length
 * 161418 165659: contig of 4242 bp in length
 * 165660 165759: gap of unknown length
 * 165760 169663: contig of 3904 bp in length
 * 169664 169763: gap of unknown length
 * 169764 173518: contig of 3755 bp in length
 * 173519 173619: gap of unknown length
 * 173619 176717: contig of 3099 bp in length
 * 176718 176818: gap of unknown length
 * 176818 179578: contig of 2761 bp in length
 * 179579 179679: gap of unknown length
 * 179679 182179: contig of 2501 bp in length
 * 182180 182279: gap of unknown length
 * 182280 184681: contig of 2402 bp in length
 * 184682 184781: gap of unknown length
 * 184782 186392: contig of 1611 bp in length
 * 186393 186493: gap of unknown length
 * 186493 188085: contig of 1592 bp in length
 * 188085 188185: gap of unknown length
 * 188185 190062: contig of 1878 bp in length
 * 190063 190163: gap of unknown length
 * 190163 191804: contig of 1641 bp in length
 * 191804 191904: gap of unknown length
 * 191904 192919: contig of 1015 bp in length
 * 192919 193019: gap of unknown length
 * 193019 194087: contig of 1068 bp in length
 * 194087 194187: gap of unknown length
 * 194187 195634: contig of 1447 bp in length
 * 195634 195734: gap of unknown length
 * 195734 196858: contig of 1125 bp in length.

FEATURES
 source

Location/Qualifiers
 1..196858
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="1"
 /clone="RP23-355F23"

BASE COUNT 55216 a 42662 c 41766 g 54102 t 3112 others
 ORIGIN

Query Match 2.4%; Score 55; DB 77; Length 196858;
 Best Local Similarity 60.3%; Pred. No. 0.0049;
 Matches 108; Conservative 0; Mismatches 70; Indels 1; Gaps 1;

QY 1258 cagtgcgaattagaagccctgagatgcctgcctgctattattagttattctctctctg 1317
 Db 176416 CACTGAGTCCAGCAACACTTGGTTGAAAAATATTCTCTTTAAAAAATATTTTTTTTTTTG 176475
 QY 1318 agacagagctcactgtgtgcccaggctagtcctcaaaacttgcggtccatttgtctcact 1377
 Db 176476 AAGAAGATCTTACTATGTAGCCCGAGGCTTGCCTTAAATGTGTGATCATCTACCTCAGT 176535
 QY 1378 catcagaatgcctgggctccagggtgtgtgcaccacactagtagctgcggttttaagct 1436
 Db 176536 CTCCTGGAGTGTGGGGTTACAGGTGTG-GTACCACAGCTGGCAGTTTCTCTTTAAAGAT 176593

Search completed: June 21, 2001, 18:32:40
 Job time: 18339 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 13:27:01 ; Search time 8015.43 Seconds
(without alignments)
4440.347 Million cell updates/sec

Title: US-09-445-201-1_COPY_8260_10560
Perfect score: 2301
Sequence: 1 tgaataagatggagggtgcc.....aattgctgaagagcctttt 2301

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues 2688314
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vil:*
59: gb_v12:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_rol:*
95: gb_rod2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2301	100.0	12845	9	AX002124	AX002124 Sequence
2	465	20.2	510	94	AF061804	AF061804 Mus muscu
3	430	18.7	430	94	AF153058	AF153058 Mus muscu
C 4	300.8	13.1	214348	66	AC021220	AC021220 Homo sapi
C 5	288.4	12.5	135552	65	AC019251	AC019251 Homo sapi
C 6	221.8	9.6	126330	63	AC013745	AC013745 Homo sapi
C 7	56.6	2.5	230352	64	AC016982	AC016982 Mus muscu
8	56	2.4	23108	94	AF088189	AF088189 Mus muscu


```

* 6525 9844: contig of 3320 bp in length
* 9845 9944: gap of unknown length
* 9945 14101: contig of 4157 bp in length
* 14102 14201: gap of unknown length
* 14202 21279: contig of 7078 bp in length
* 21280 21379: gap of unknown length
* 21380 31346: contig of 9967 bp in length
* 31347 31447: gap of unknown length
* 31448 40526: contig of 9080 bp in length
* 40527 40626: gap of unknown length
* 40627 52862: contig of 12236 bp in length
* 52863 52962: gap of unknown length
* 52963 66090: contig of 13128 bp in length
* 66091 66191: gap of unknown length
* 66192 82759: contig of 16469 bp in length
* 82760 82660: gap of unknown length
* 82661 104797: contig of 22038 bp in length
* 104798 104898: gap of unknown length
* 104899 129319: contig of 24422 bp in length
* 129320 129420: gap of unknown length
* 129421 176511: contig of 47092 bp in length
* 176512 176612: gap of unknown length
* 176613 177666: contig of 1055 bp in length
* 177667 177767: gap of unknown length
* 177768 179097: contig of 1231 bp in length
* 179098 180242: contig of unknown length
* 180243 180343: gap of unknown length
* 180344 181442: contig of 1100 bp in length
* 181443 181542: gap of unknown length
* 181543 182642: contig of 1100 bp in length
* 182643 182742: gap of unknown length
* 182743 184165: contig of 1423 bp in length
* 184166 184265: gap of unknown length
* 184266 185512: contig of 1247 bp in length
* 185513 185612: gap of unknown length
* 185613 187121: contig of 1509 bp in length
* 187122 187221: gap of unknown length
* 187222 188524: contig of 1303 bp in length
* 188525 188624: gap of unknown length
* 188625 189782: contig of 1158 bp in length
* 189783 189882: gap of unknown length
* 189883 191312: contig of 1430 bp in length
* 191313 191412: gap of unknown length
* 191413 192837: contig of 1425 bp in length
* 192838 192937: gap of unknown length
* 192938 194117: contig of 1180 bp in length
* 194118 194217: gap of unknown length
* 194218 195948: contig of 1731 bp in length
* 195949 196048: gap of unknown length
* 196049 197744: contig of 1696 bp in length
* 197745 197844: gap of unknown length
* 197845 199489: contig of 1645 bp in length
* 199490 199589: gap of unknown length
* 199590 201304: contig of 1715 bp in length
* 201305 201404: gap of unknown length
* 201405 203316: contig of 1912 bp in length
* 203317 203416: gap of unknown length
* 203417 204490: contig of 1074 bp in length
* 204491 204590: gap of unknown length
* 204591 206676: contig of 2086 bp in length
* 206677 206776: gap of unknown length
* 206777 208803: contig of 2027 bp in length
* 208804 208903: gap of unknown length
* 208904 210597: contig of 1694 bp in length
* 210598 210697: gap of unknown length
* 210698 212097: contig of 1400 bp in length
* 212098 212197: gap of unknown length
* 212198 214348: contig of 2151 bp in length.
FEATURES
  Location/Qualifiers
    1..214348
      /organism="Homo sapiens"
      /db_xref="taxon:9606"

```

source

Query Match

13.1%; Score 300.8; DB 66; Length 214348;

```

/chromosome="4"
1..1584
  /note="assembly_name:Contig100"
1685..4194
  /note="assembly_name:Contig101"
4295..6424
  /note="assembly_name:Contig102"
6525..9844
  /note="assembly_name:Contig103"
9945..14101
  /note="assembly_name:Contig104"
14202..21279
  /note="assembly_name:Contig105"
21380..31346
  /note="assembly_name:Contig106"
31447..40526
  /note="assembly_name:Contig107"
40627..52862
  /note="assembly_name:Contig108"
52963..66090
  /note="assembly_name:Contig109"
66191..82659
  /note="assembly_name:Contig110"
82760..104797
  /note="assembly_name:Contig111"
104898..129319
  /note="assembly_name:Contig112"
129420..176511
  /note="assembly_name:Contig113"
176612..177666
  /note="assembly_name:Contig133"
177767..178997
  /note="assembly_name:Contig137"
179098..180242
  /note="assembly_name:Contig156"
180343..181442
  /note="assembly_name:Contig159"
181543..182642
  /note="assembly_name:Contig166"
182743..184165
  /note="assembly_name:Contig169"
184266..185512
  /note="assembly_name:Contig175"
185613..187121
  /note="assembly_name:Contig177"
187222..188524
  /note="assembly_name:Contig178"
188625..189782
  /note="assembly_name:Contig180"
189883..191312
  /note="assembly_name:Contig185"
191413..192837
  /note="assembly_name:Contig186"
192938..194117
  /note="assembly_name:Contig187"
194218..195948
  /note="assembly_name:Contig188"
196049..197744
  /note="assembly_name:Contig189"
197845..199489
  /note="assembly_name:Contig190"
199590..201304
  /note="assembly_name:Contig191"
201405..203316
  /note="assembly_name:Contig193"
203417..204490

```

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

*	18310	18409:	gap of	100	bp
*	18410	19183:	contig of	774	bp in length
*	19184	19283:	gap of	100	bp
*	19284	20038:	contig of	755	bp in length
*	20039	20138:	gap of	100	bp
*	20139	20878:	contig of	740	bp in length
*	20879	20978:	gap of	100	bp
*	20979	21780:	contig of	802	bp in length
*	21781	21880:	gap of	100	bp
*	21881	22642:	contig of	762	bp in length
*	22643	22742:	gap of	100	bp
*	22743	23549:	contig of	807	bp in length
*	23550	23649:	gap of	100	bp
*	23650	24442:	contig of	793	bp in length
*	24443	24542:	gap of	100	bp
*	24543	25313:	contig of	771	bp in length
*	25314	25413:	gap of	100	bp
*	25414	26164:	contig of	751	bp in length
*	26165	26264:	gap of	100	bp
*	26265	27045:	contig of	781	bp in length
*	27046	27145:	gap of	100	bp
*	27146	27921:	contig of	776	bp in length
*	27922	28021:	gap of	100	bp
*	28022	28812:	contig of	791	bp in length
*	28813	28912:	gap of	100	bp
*	28913	29663:	contig of	751	bp in length
*	29664	29763:	gap of	100	bp
*	29764	30555:	contig of	792	bp in length
*	30556	30655:	gap of	100	bp
*	30656	31427:	contig of	772	bp in length
*	31428	31527:	gap of	100	bp
*	31528	32317:	contig of	790	bp in length
*	32318	32417:	gap of	100	bp
*	32418	33198:	contig of	781	bp in length
*	33199	33298:	gap of	100	bp
*	33299	34081:	contig of	783	bp in length
*	34082	34181:	gap of	100	bp
*	34182	34938:	contig of	757	bp in length
*	34939	35038:	gap of	100	bp
*	35039	35795:	contig of	757	bp in length
*	35796	35895:	gap of	100	bp
*	35896	36624:	contig of	729	bp in length
*	36625	36724:	gap of	100	bp
*	36725	37501:	contig of	777	bp in length
*	37502	37601:	gap of	100	bp
*	37602	38375:	contig of	774	bp in length
*	38376	38475:	gap of	100	bp
*	38476	39264:	contig of	789	bp in length
*	39265	39364:	gap of	100	bp
*	39365	40148:	contig of	784	bp in length
*	40149	40248:	gap of	100	bp
*	40249	41001:	contig of	753	bp in length
*	41002	41101:	gap of	100	bp
*	41102	41871:	contig of	770	bp in length
*	41872	41971:	gap of	100	bp
*	41972	42757:	contig of	786	bp in length
*	42758	42857:	gap of	100	bp
*	42858	43646:	contig of	789	bp in length
*	43647	43746:	gap of	100	bp
*	43747	44507:	contig of	761	bp in length
*	44508	44607:	gap of	100	bp
*	44608	45362:	contig of	755	bp in length
*	45363	45462:	gap of	100	bp
*	45463	46193:	contig of	731	bp in length
*	46194	46293:	gap of	100	bp
*	46294	47090:	contig of	797	bp in length
*	47091	47190:	gap of	100	bp
*	47191	47981:	contig of	791	bp in length
*	47982	48081:	gap of	100	bp
*	48082	48867:	contig of	786	bp in length
*	48868	48967:	gap of	100	bp
*	48968	49724:	contig of	757	bp in length
*	49725	49824:	gap of	100	bp

*	49825	50586:	contig of 762 bp in length
*	50587	50686:	gap of 100 bp
*	50687	51456:	contig of 770 bp in length
*	51457	51556:	gap of 100 bp
*	51557	52334:	contig of 778 bp in length
*	52335	52434:	gap of 100 bp
*	52435	53204:	contig of 770 bp in length
*	53205	53304:	gap of 100 bp
*	53305	54077:	contig of 773 bp in length
*	54078	54177:	gap of 100 bp
*	54178	54958:	contig of 781 bp in length
*	54959	55058:	gap of 100 bp
*	55059	55833:	contig of 775 bp in length
*	55834	55933:	gap of 100 bp
*	55934	56716:	contig of 783 bp in length
*	56717	56816:	gap of 100 bp
*	56817	57612:	contig of 796 bp in length
*	57613	57712:	gap of 100 bp
*	57713	58422:	contig of 710 bp in length
*	58423	58522:	gap of 100 bp
*	58523	59310:	contig of 788 bp in length
*	59311	59410:	gap of 100 bp
*	59411	60186:	contig of 776 bp in length
*	60187	60286:	gap of 100 bp
*	60287	61061:	contig of 775 bp in length
*	61062	61161:	gap of 100 bp
*	61162	61891:	contig of 730 bp in length

Query Match 12.5%; Score 288.4; DB 65; Length 135552;
Best Local Similarity 77.5%; Pred. No. 1.2e-60;
Matches 406; Conservative 0; Mismatches 101; Indels 17; G

Qy	1793	gaactggagagatat	tttttcttccccaat	tcctttctccatt	taaatgctgtc	cttt	1852	
Db	63618	GAATTGGGAATG	TTTTTCTCTCTCCCGT	TCTCTCCCTTG	CTTTTAATGTACAGT	CTTT	63559	
Qy	1853	agaagcaactgcctc	agcttctgaagctc	agatacccaagaag	actctggtacacacg	atg	1912	
Db	63558	ATGAGCCATTATT	CAGCTGTGGCAGT	TGTGTATCACGGGAAG	CGCACTAGA-AAAT	TG	63500	
Qy	1913	ataaagacaatgg	agcgggtccag	tggtcccg	tcctctcaggg	tatggagacga	1972	
Db	63499	ATAAAGGAAAT	GAGACAAGGT	CATAGATTCTCTC	ACTCCCTTCA-GGT	ACGTAGATGA	63441	
Qy	1973	gctgtagagagatg	tctccagggag	tttcttaatacag	caatttagtcagatc	gtgc	2032	
Db	63440	ACTATATAAAA	TCGTAACTGGAT	TGCTTATATCAG	CAATTTAGTCAAA	TGTGTACA	63381	
Qy	2033	tcctatgctttaca	agaaatgcag	tggggcc-----	tgagatcatc	agatg	2078	
Db	63380	TCCATATGTTCT	ATAAGAAATG	TCAGTGGGTCT	TTTCCCAAGSGAG	TGAGATCATCAGATG	63321	
Qy	2079	gaggttcac	cggtttc	aatgtcccg	tatccttttg	taagaccttgaa	gtgcgaacgca	2138
Db	63320	AAGGTTCAAT	TGTTTTC	AATGTCCCGT	TATCCTTTTGT	AAGACCTTGAAG	TGTGGCAATGCA	63261
Qy	2139	ggaaaacagga	actccacctgg	tcgctg	gaattgcag	actgtgtgtgtg	ttgtgtgac	2198
Db	63260	GGAAACAGG	AATCCACCT	TAGCTCCAT	GAAATTCACAG	ACTGTGTGTGTGT	TATGTATGAC	63201
Qy	2199	catctgcccattc	tctcgtgttat	gacagagctgt	gaaacttta	actggagctggg	ggcaaa	2258
Db	63200	CATCTGCCCAT	TCTTCCTGTT	TATGACACAG	CTTGTGAACTTT	ACTGAGAA	TGGTGAAA	63141
Qy	2259	gtcaa-tccc	actttataca	atgaattg	ctggaagag	ggcctttt	2301	
Db	63140	GTAATATCCC	AGTTTTTAT	TACAAATGA	ATTAATGCT	CAAGAGGC	CTTTT	63097

RESULT	6
AC013745/C	
LOCUS	
DEFINITION	

```

ACCESSION      AC013745
VERSION        AC013745.3
KEYWORDS       HTG; HTGS_PHASE0
SOURCE         human.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 126330)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 4, clone RP11-92P18
unpublished
2 (bases 1 to 126330)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (15-NOV-1999), Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6910837.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2025
Center clone name: 92_P_18
-----
* NOTE: This record contains 147 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 723 822: contig of 722 bp in length
* 823 822: gap of 100 bp
* 823 1542: contig of 720 bp in length
* 1543 1642: gap of 100 bp
* 1643 2371: contig of 729 bp in length
* 2372 2471: gap of 100 bp
* 2472 3212: contig of 741 bp in length
* 3213 3312: gap of 100 bp
* 3313 4049: contig of 737 bp in length
* 4050 4149: gap of 100 bp
* 4150 4840: contig of 691 bp in length
* 4841 4940: gap of 100 bp
* 4941 5676: contig of 736 bp in length
* 5677 5776: gap of 100 bp
* 5777 6521: contig of 745 bp in length
* 6522 6621: gap of 100 bp
* 6622 7344: contig of 723 bp in length
* 7345 7444: gap of 100 bp
* 7445 8161: contig of 717 bp in length
*
* 8162 8261: gap of 100 bp
* 8262 9014: contig of 753 bp in length
* 9015 9114: gap of 100 bp
* 9115 9827: contig of 713 bp in length
* 9828 9927: gap of 100 bp
* 9928 10650: contig of 723 bp in length
* 10651 10750: gap of 100 bp
* 10751 11479: contig of 729 bp in length
* 11480 11579: gap of 100 bp
* 11580 12318: contig of 739 bp in length
* 12319 12418: gap of 100 bp
* 12419 13137: contig of 719 bp in length
* 13138 13237: gap of 100 bp
* 13238 13259: contig of 722 bp in length
* 13259 14059: gap of 100 bp
* 14060 14785: contig of 726 bp in length
* 14786 14885: gap of 100 bp
* 14886 15580: contig of 695 bp in length
* 15581 15680: gap of 100 bp
* 15681 16420: contig of 740 bp in length
* 16421 16520: gap of 100 bp
* 16521 17233: contig of 713 bp in length
* 17234 17333: gap of 100 bp
* 17334 18057: contig of 724 bp in length
* 18058 18157: gap of 100 bp
* 18158 18874: contig of 717 bp in length
* 18875 18974: gap of 100 bp
* 18975 19716: contig of 742 bp in length
* 19717 19816: gap of 100 bp
* 19817 20557: contig of 741 bp in length
* 20558 20657: gap of 100 bp
* 20658 21379: contig of 722 bp in length
* 21380 21479: gap of 100 bp
* 21480 22193: contig of 714 bp in length
* 22194 22293: gap of 100 bp
* 22294 23013: contig of 720 bp in length
* 23014 23113: gap of 100 bp
* 23114 23851: contig of 738 bp in length
* 23852 23951: gap of 100 bp
* 23952 24681: contig of 730 bp in length
* 24682 24781: gap of 100 bp
* 24782 25505: contig of 724 bp in length
* 25506 25605: gap of 100 bp
* 25606 26335: contig of 730 bp in length
* 26336 26435: gap of 100 bp
* 26436 27158: contig of 723 bp in length
* 27159 27258: gap of 100 bp
* 27259 27960: contig of 702 bp in length
* 27961 28060: gap of 100 bp
* 28061 28775: contig of 715 bp in length
* 28776 28875: gap of 100 bp
* 28876 29586: contig of 711 bp in length
* 29587 29686: gap of 100 bp
* 29687 30394: contig of 708 bp in length
* 30395 30494: gap of 100 bp
* 30495 31246: contig of 752 bp in length
* 31247 31346: gap of 100 bp
* 31347 32090: contig of 744 bp in length
* 32091 32190: gap of 100 bp
* 32191 32894: contig of 704 bp in length
* 32895 32994: gap of 100 bp
* 32995 33741: contig of 747 bp in length
* 33742 33841: gap of 100 bp
* 33842 34602: contig of 761 bp in length
* 34603 34702: gap of 100 bp
* 34703 35440: contig of 738 bp in length
* 35441 35540: gap of 100 bp
* 35541 36261: contig of 721 bp in length
* 36262 36361: gap of 100 bp
* 36362 37081: contig of 720 bp in length
* 37082 37181: gap of 100 bp
* 37182 37898: contig of 717 bp in length
* 37899 37998: gap of 100 bp

```

```

* 37999 38730: contig of 732 bp in length
* 38731 38830: gap of 100 bp
* 38831 39526: contig of 696 bp in length
* 39527 39626: gap of 100 bp
* 39627 40316: contig of 690 bp in length
* 40317 40416: gap of 100 bp
* 40417 41136: contig of 720 bp in length
* 41137 41236: gap of 100 bp
* 41237 41985: contig of 749 bp in length
* 41986 42085: gap of 100 bp
* 42086 42815: contig of 730 bp in length
* 42816 42915: gap of 100 bp
* 42916 43632: contig of 717 bp in length
* 43633 43732: gap of 100 bp
* 43733 44445: contig of 713 bp in length
* 44446 44545: gap of 100 bp
* 44546 45287: contig of 742 bp in length
* 45288 45387: gap of 100 bp
* 45388 46129: contig of 742 bp in length
* 46130 46229: gap of 100 bp
* 46230 46961: contig of 732 bp in length
* 46962 47061: gap of 100 bp
* 47062 47765: contig of 704 bp in length
* 47766 47865: gap of 100 bp
* 47866 48582: contig of 717 bp in length
* 48583 48682: gap of 100 bp
* 48683 49429: contig of 747 bp in length
* 49430 49529: gap of 100 bp
* 49530 50236: contig of 707 bp in length
* 50237 50336: gap of 100 bp
* 50337 51082: contig of 726 bp in length
* 51063 51162: gap of 100 bp
* 51163 51884: contig of 722 bp in length
* 51885 51984: gap of 100 bp
* 51985 52702: contig of 718 bp in length
* 52703 52802: gap of 100 bp
* 52803 53516: contig of 714 bp in length
* 53517 53616: gap of 100 bp
* 53617 54329: contig of 713 bp in length
* 54330 54429: gap of 100 bp
* 54430 55140: contig of 711 bp in length
* 55141 55240: gap of 100 bp
* 55241 55930: contig of 690 bp in length
* 55931 56030: gap of 100 bp
* 56031 56724: contig of 694 bp in length
* 56725 56824: gap of 100 bp
* 56825 57628: contig of 804 bp in length
* 57629 57728: gap of 100 bp
* 57729 58523: contig of 795 bp in length
* 58524 58623: gap of 100 bp
* 58624 59387: contig of 764 bp in length
* 59388 59487: gap of 100 bp

Query Match          9.6%; Score 221.8; DB 63; Length 126330;
Best Local Similarity 64.8%; Pred. No. 3.5e-44;
Matches 276; Conservative 0; Mismatches 135; Indels 15; Gaps 2;

QY 1891 aggaagtctgtacacagcatgataaagacacatggagcgggtcacagtggctccgcgc 1950
| | | | |
Db 88902 ANNACAAAGTTATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 88843
| | | | |
QY 1951 ccttcaggggatgagacagcgtgtagagagatgtctccaggaggttcttaataca 2010
| | | | |
Db 88842 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 88783
| | | | |
QY 2011 gcaatttagcagatctgtcctcctatgctttacagaagaatgtcagtgagcc----- 2063
| | | | |
Db 88782 GCAATTAGTCAATGTGTACATCCCTATGTCTTATAGAAATGTCAATGGTCCCTTTCC 88723
| | | | |
QY 2064 -----tagatcatcagatcaggttcctcgtttcattcattcattcattcatttga 2116
| | | | |
Db 88722 AAGGGAGTGAATCATCAGATGAAGGTTCAATTTGGTTTCAATGTCCTGATCCTTTGTA 88663
| | | | |

```

```

QY 2117 agacctgaagtgtgcaacgcagagaaacaggaactccacccctgggtgcggaattgcag 2176
| | | | |
Db 88662 AGACCTTGAAGTTGGCAATGCAGAAACAGAACTCACCTAGCTCCATGAATTGCAG 88603
| | | | |
QY 2177 agctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2236
| | | | |
Db 88602 AACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 88543
| | | | |
QY 2237 ctttaactgggactggggcacaagtcaa-tcccacactttatacaatgaattgtcgaaggg 2295
| | | | |
Db 88542 CTTTACTGAGAAAGTGGTGAAGAACTAAATCCCACTTTTATACAATGAATTCCTGAAGAG 88483
| | | | |
QY 2296 cctttt 2301
| | | | |
Db 88482 CCTTTT 88477
| | | | |

RESULT 7
AC016982/c
LOCUS
DEFINITION AC016982 230352 bp DNA HTG 26-FEB-2001
MUS musculus clone RP23-416H2, WORKING DRAFT SEQUENCE, 11 unordered
pieces.
ACCESSION AC016982
VERSION AC016982.4 GI:13123354
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 230352)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Mus musculus, clone RP23-416H2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 230352)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Becker, R., Bedalov, F.,
Boguslavskiy, L., Bouckgalter, B., Brown, A., Castle, A., Colangelo, M.,
Collins, S., Collumore, A., Cooke, P., Dearellano, K., Dewar, K.,
Domino, M., Doyle, M., Fenestor, J., Ferreira, P., Fitzhugh, W.,
Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B.,
Heaford, A., Horton, L., Howland, J.C., Johnson, R., Jones, C., Kann, L.,
Karatas, A., Klein, J., Landers, T., Lechoczky, J., Lieu, C., Locke, K.,
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
Meldrum, J., Morrow, J., Naylor, J., Norman, C.H., O'Connor, T.,
O'Donnell, P., Peterson, K., Pierre, N., Pollara, V., Riley, R.,
Rothman, D., Roy, A., Santos, R., Severy, P., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
Zimmer, A. and Zody, M.
Direct Submission
Submitted (09-DEC-1999), Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 25, 2001 this sequence version replaced gi:8072589.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5038
Center clone name: 416_H_2
----- Summary Statistics
Sequencing vector: M13; M7815; 39% of reads
Sequencing vector: Plasmid; n/a; 61% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 226546 bases at least Q40
Consensus quality: 228022 bases at least Q30
Consensus quality: 228658 bases at least Q20
Insert size: 212000; agarose-fp

```

Insert size: 229352; sum-of-contigs
Quality coverage: 9.5 in Q20 bases; agarose-fp
Quality coverage: 8.8 in Q20 ba.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 4159: contig of 4159 bp in length
* 4160 4259: gap of 100 bp
* 4260 6386: contig of 2127 bp in length
* 6387 6486: gap of 100 bp
* 6487 59418: contig of 52932 bp in length
* 59419 59518: gap of 100 bp
* 59519 69698: contig of 10180 bp in length
* 69699 69798: gap of 100 bp
* 69799 83694: contig of 13896 bp in length
* 83695 83794: gap of 100 bp
* 83795 101615: contig of 17821 bp in length
* 101616 101715: gap of 100 bp
* 101716 127243: contig of 25528 bp in length
* 127244 127343: gap of 100 bp
* 127344 158706: contig of 31363 bp in length
* 158707 158806: gap of 100 bp
* 158807 191011: contig of 32205 bp in length
* 191012 191111: gap of 100 bp
* 191112 222984: contig of 31873 bp in length
* 222985 223084: gap of 100 bp
* 223085 230352: contig of 7268 bp in length.

FEATURES

source
1. .230352
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-416H2"
/clone_lib="RPCI-23 Female Mouse BAC"
misc_feature
1. .4159
/note="assembly_fragment
clone_end:SP6
vector_side:left"
misc_feature
4260. .6386
/note="assembly_fragment"
misc_feature
6487. .59418
/note="assembly_fragment"
misc_feature
59519. .69698
/note="assembly_fragment"
misc_feature
69799. .83694
/note="assembly_fragment"
misc_feature
83795. .101615
/note="assembly_fragment"
misc_feature
101716. .127243
/note="assembly_fragment"
misc_feature
127344. .158706
/note="assembly_fragment"
misc_feature
158807. .191011
/note="assembly_fragment"
misc_feature
191112. .222984
/note="assembly_fragment"
misc_feature
223085. .230352
/note="assembly_fragment
clone_end:T7
vector_side:right"
BASE COUNT 60350 a 54120 c 54747 g 60133 t 1002 others
ORIGIN

Query Match 2.5%; Score 56.6; DB 64; Length 230352;
Best Local Similarity 71.8%; Pred. No. 0.0025;
Matches 74; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 1381 ttcttcttctgagacagagtctactgtgtgcccagagtagtctcaaaacttgcggtc 1440

Db 110701 TTTTCTCTTTTGTGAGACAGTCTCACTGTGTAAACCCAGGCTGCTCTCAAGCTGTGGCC 110642
QY 1441 catttgttcacatcacagaatgctggcttcagatgtgtgc 1483
Db 110641 CCTCTGTTTTCGTCCTCCCAAGTCTGGGATTACAAGTGTGTAC 110599

RESULT 8

AF088189 AF088189 23108 bp DNA ROD 28-FEB-1999
LOCUS Mus musculus CD4 antigen (Cd4) gene, partial sequence.
DEFINITION AF088189
ACCESSION AF088189
VERSION AF088189.1 GI:4322490
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 23108)
AUTHORS Hansen,R.K., McCreedy,P.M. and Sands,J.F.
TITLE Sequence Analysis of the 5' Flanking Region and First Intron of the
MURINE CD4 Gene
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 23108)
AUTHORS Hansen,R.K., McCreedy,P.M. and Sands,J.F.
TITLE Direct Submission
JOURNAL Submitted (31-AUG-1998) Biochemistry, Loma Linda, MT219, Loma
Linda, CA 92354, USA
FEATURES
source
1. .23108
/organism="Mus musculus"
/strain="B10.D2/nSnJ"
/db_xref="taxon:10090"
/chromosome="6"
join(14561. .14699,23082. .>23108)
/gene="Cd4"
/product="CD4 antigen"
join(14561. .14699,23082. .>23108)
/gene="Cd4"
14561. .14699
/gene="Cd4"
/number=1
14561. .>23108
/gene="Cd4"
14564
misc_feature
/gene="Cd4"
misc_feature
/note="alternate transcription start site"
14569
/gene="Cd4"
/note="alternate transcription start site"
17927. .18271
/gene="Cd4"
/note="first intron silencer; DNase I hypersensitivity
site 10 (DH10)"
23082. .>23108
/gene="Cd4"
/number=2
BASE COUNT 6086 a 5263 c 5514 g 6245 t
ORIGIN

Query Match 2.4%; Score 56; DB 94; Length 23108;
Best Local Similarity 58.3%; Pred. No. 0.0034;
Matches 98; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 1371 taattagttattctcttctgagacagagtcactgtgtgcccaggtagtctcaa 1430
Db 3026 TATTATTATTATTCATTATTATTGGAGCAGTGTCTAACTATGTAGTACAGCTGCCTACA 3085
QY 1431 acttgcggtccattgtctcactcatcagaatgctgggtccaggtgtgtgtgcaccacac 1490
Db 3086 ACTCTCTTTCCTGTAGTCTGACCCCTCCAGGACTGGGGTTACAGGTGTGAGGGCACAAC 3145


```
/translation="MPESYFNCPDGEYQSNDCVCKTCPSGTFFVKAPCKIPHTQGOCE
KCHPFTGDNGLHDELCSTCDQNMWADCSATSDRKCCEQLIYYIDPKFPESC
RPTCKCPGIPVLQECNCTANTVCSYSSVSNPRNWLFLMLIVFCI"
37049..37176
/gene="Tnfrh1"
/number=2
46179..46291
/gene="Tnfrh1"
/number=3
47068..47141
/gene="Tnfrh1"
/number=4
48531..48580
/gene="Tnfrh1"
/number=5
join(<67525..67596,72244..72334,73711..73838,76251..76363,
77140..77213,77826..77868,78677..78828)
/gene="Tnfrh2"
join(67525..67596,72244..72334,73711..73838,76251..76363,
77140..77213,77826..77868,78677..78828)
/gene="Tnfrh2"
join(72256..72334,73711..73838,76251..76363,77140..77213,
77826..77857)

Query Match      2.4%  Score 56; DB 94; Length 281000;
Best Local Similarity 54.3%; Pred. No. 0.0035;
Matches 113; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 1403 ctcaactgtgcccagctagctcacaactgcggtccattgtctcactcatcagaat 1462
||||| ||| ||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 275809 CTCACATGATGAGCCAGGACGCTCTAGCTTAAAGTCTCTTGCTCAGCCTCTTAAT 275868

QY 1463 gctgggttcacagtggtgtgcacacacactaggtagctgcgttttaagctgaagctgga 1522
||||| ||||| ||||| ||| || ||||| ||||| ||||| ||||| ||||| |||||
Db 275869 GCTGGGATTCAGGATGATATATCCGGCAGCTCTCTGCAACTGTTCAAACTGGAAATCTTTA 275928

QY 1523 agatccctgatctcttaccatgttggtgcacacactaggttagctgactgaaactagtt 1582
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 275929 ACCAATAGATGTTTAAAGAGCCCAATAGAGATGAGTGTGTAGAAACAGACATTCGAA 275988

QY 1583 atctcgtgtgtaataacacgtcagtggt 1610
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 275989 ATGTGGTCAATAAAGCACCTAACCGT 276016

RESULT 13
AC073787
LOCUS
DEFINITION
AC073787 166513 bp DNA HTG 18-JUL-2000
Mus musculus clone RP23-413L18, WORKING DRAFT SEQUENCE, 16 ordered
pieces.
AC073787
VERSION
AC073787.2 GI:9256796
KEYWORDS
HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 166513)
DOE Joint Genome Institute.
Sequencing of Mouse
Unpublished
REFERENCE
2 (bases 1 to 166513)
DOE Joint Genome Institute.
Direct Submission
Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:8810404.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
```

Center Project Name: 1895746
Center clone name: RPCI-23_413L18

Summary Statistics
Consensus quality: 157635 bases at least Q40
Consensus quality: 163578 bases at least Q30
Consensus quality: 164611 bases at least Q20
Estimated insert size: 183000; agarose-fp estimation
Quality coverage: 6.23 in Q20 bases; agarose-fp estimation
Quality coverage: 6.88 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 conligs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 2215 2214: contig of 2214 bp in length
2315 15749: contig of 13435 bp in length
15850 15849: gap of unknown length
31750 31759: contig of 15910 bp in length
31760 31859: gap of unknown length
31860 47298: contig of 15439 bp in length
47299 47398: gap of unknown length
47399 50191: contig of 2793 bp in length
50192 50291: gap of unknown length
50292 53149: contig of 2858 bp in length
53150 53249: gap of unknown length
53250 64766: contig of 11517 bp in length
64767 64866: gap of unknown length
64867 67070: contig of 2204 bp in length
67071 67170: gap of unknown length
67171 87473: contig of 20303 bp in length
87474 87573: gap of unknown length
87574 97682: contig of 10109 bp in length
97683 97782: gap of unknown length
97783 115608: contig of 17826 bp in length
115609 115708: gap of unknown length
115709 127399: contig of 11691 bp in length
127400 127499: gap of unknown length
127500 128951: contig of 1452 bp in length
128952 129051: gap of unknown length
129052 149899: contig of 20848 bp in length
149900 149999: gap of unknown length
150000 151700: contig of 1701 bp in length
151701 151800: gap of unknown length
151801 166513: contig of 14713 bp in length.

FEATURES

Location/Qualifiers

1..166513
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-413L18"

BASE COUNT 39178 a 43350 c 43117 g 39368 t 1500 others
ORIGIN

Query Match

Best Local Similarity 58.0%; Pred. No. 0.0039;
Matches 120; Conservative 0; Mismatches 82; Indels 5; Gaps 1;

QY 1361 tgctgtctataatagttattctctctctgagacagagctcactgtgtggccag 1420
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 82983 TGAGTGCACATTTCTTCTTTTGTTCCTGAGACAGACTCCCTATGTACCTAGG 83042
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1421 ctagtctcaaaccttggtccattgtctcactcatcagaatgtgggtccaggtgtg 1480
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 83043 CTGGCCTCCAACTAGTGTCTCTGCTCAGCCTTCCAAATGCTGGATTACAAATATG 83102
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 1481 tgcaccacactagtagtcgcgttttaagctaagagctggaagactcgtatgctcttta 1540
 Db 83103 TG-----CCTTAGCTGGCCCAAGGTATAAATCTACAGTGGCAGCTGCTAATCTTTTGA 83157
 Qy 1541 ccattggtggcattgtacaggttaagt 1567
 Db 83158 AGTAGGGGGTGGGTGTTCTTTGTTGTT 83184

RESULT 14

AC019026 LOCUS 157996 bp DNA ROD 14-JUL-2000
 DEFINITION Mus musculus chromosome 6 clone RP23-188E5 strain C57BL6/J,
 complete sequence.
 AC019026
 VERSION AC019026.12 GI:8927595
 KEYWORDS HTG.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 157996)
 AUTHORS Montgomery K.T., Grills G., Han J., Lee E., Long J., Pomerantz R.,
 Toshikhes I.P., Shim C., Decker J., Thomas E., Perera A. and
 Kucheralapati R.
 TITLE Mouse High Throughput Sequencing
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 157996)
 AUTHORS Montgomery K.T., Grills G., Han J., Lee E., Long J., Pomerantz R.,
 Toshikhes I.P., Shim C., Decker J., Thomas E., Perera A. and
 Kucheralapati R.
 TITLE Direct Submission
 JOURNAL Submitted (30-DEC-1999) Department of Molecular Genetics, Albert
 Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
 Bronx, NY 10461, USA
 REFERENCE 3 (bases 1 to 157996)
 AUTHORS Montgomery K.T., Grills G., Han J., Lee E., Long J., Pomerantz R.,
 Toshikhes I.P., Shim C., Decker J., Thomas E., Perera A. and
 Kucheralapati R.
 TITLE Direct Submission
 JOURNAL Submitted (06-JUL-2000) Department of Molecular Genetics, Albert
 Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
 Bronx, NY 10461, USA
 REFERENCE 4 (bases 1 to 157996)
 AUTHORS Montgomery K.T., Grills G., Han J., Lee E., Long J., Pomerantz R.,
 Toshikhes I.P., Shim C., Decker J., Thomas E., Perera A. and
 Kucheralapati R.
 TITLE Direct Submission
 JOURNAL Submitted (14-JUL-2000) Department of Molecular Genetics, Albert
 Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
 Bronx, NY 10461, USA

COMMENT

On Jul 6, 2000 this sequence version replaced gi:7767645.
 -----Genome Center Center:
 Albert Einstein College of Medicine Center
 Code: AECOM
 Web site:
<http://sequence.aecom.yu.edu/cgi-bin/ws.exe/mouseDB/mouseSEQ/mouseseqtable.htm>
 Contact: jhan@sequence.aecom.yu.edu

CLONE LENGTH: This sequence represents the entire insert of this
 clone unless otherwise noted. If there are overlapping clones, the
 overlaps are noted in the beginning and end of the Features
 listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550).
 Repeats are identified using RepeatMasker (A. Smit and P. Green,
 unpublished.) for Human and Mouse sequences.
 Genes and Regions of sequence similarity are identified by BLAST
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to EST
 and cDNA sequences in Unigene. Genes demonstrate at least two exons
 flanked by consensus splice sites that maintain sequence continuity

across the splice junctions. Sequences that are not identical
 matches are annotated as similar.

SEQUENCING READ COVERAGE: Attempts are made to complete double
 stranded sequence for all regions. All sequence is completed to a
 standard of coverage with a minimum of 3 reads with no ambiguities.
 If the sequence coverage for a region does not meet this standard,
 it is indicated in the annotation as Low Coverage. Low coverage
 linkages are verified by PCR product size verification or
 verification of forward and reverse reads from clones which span
 the low coverage area.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
 standards - estimated average error rate is less than 1 per 10,000
 bases using the Consed quality parameters. Regions that do not
 meet this requirement are annotated as Low Quality.

-----Summary Statistics:

Center project name: AAP
 Sequencing vector: pUC18; L08752
 Chemistry: Dye-terminator Big Dye; 100%
 Assembly program: Phrap version 0.990319
 Contig length: 157996
 Fraction of Phrap value < 40: 0.0011076
 Error rate shown by Consed: 0.13 per 10,000 bases
 Number of N's in consensus: 1

----- Distribution of Quality < 40 Bases:

 1000
 900
 800
 700
 600
 500
 400
 300
 200
 100
 0
 5 10 15 20 25 30 35 40
 Phrap Value Range

FEATURES	source	Location/Qualifiers
repeat_region	1..157996	/organism="Mus musculus"
repeat_region	1..157996	/db_xref="taxon:10090"
repeat_region	1..157996	/chromosome="6"
repeat_region	1..157996	/clone="RP23-188E5"
repeat_region	1..157996	448..743
repeat_region	1..157996	/rpt_family="(TG)n"
repeat_region	1..157996	complement(2116..2138)
repeat_region	1..157996	/rpt_family="ID3"
repeat_region	1..157996	complement(2139..2274)
repeat_region	1..157996	/rpt_family="B1_MM"
repeat_region	1..157996	complement(2275..2348)
repeat_region	1..157996	/rpt_family="ID3"
repeat_region	1..157996	complement(3344..3682)
repeat_region	1..157996	/rpt_family="I1_MM"
repeat_region	1..157996	4251..4347
repeat_region	1..157996	/rpt_family="CT-rich"
repeat_region	1..157996	4348..4459
repeat_region	1..157996	/rpt_family="(CATA)n"
repeat_region	1..157996	4631..4754
repeat_region	1..157996	/rpt_family="B4A"
repeat_region	1..157996	4807..4841
repeat_region	1..157996	/rpt_family="(CATA)n"
repeat_region	1..157996	4863..4884
repeat_region	1..157996	/rpt_family="AT-rich"
repeat_region	1..157996	5089..5164
repeat_region	1..157996	/rpt_family="AT-rich"
repeat_region	1..157996	5220..5293

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 14:16:32 ; Search time 346.98 Seconds
(without alignments)
3871.349 Million cell updates/sec

Title: US-09-445-201-1_COPY_8260_10560
Perfect score: 2301
Sequence: 1 tgaataagatgaggtgccc.....aattgctgaagagcctttt 2301

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues
Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0401:*

- 1: /SID22/gcgdata/geneseq/geneseq/NA1980.DAT:*
- 2: /SID22/gcgdata/geneseq/geneseq/NA1981.DAT:*
- 3: /SID22/gcgdata/geneseq/geneseq/NA1982.DAT:*
- 4: /SID22/gcgdata/geneseq/geneseq/NA1983.DAT:*
- 5: /SID22/gcgdata/geneseq/geneseq/NA1984.DAT:*
- 6: /SID22/gcgdata/geneseq/geneseq/NA1985.DAT:*
- 7: /SID22/gcgdata/geneseq/geneseq/NA1986.DAT:*
- 8: /SID22/gcgdata/geneseq/geneseq/NA1987.DAT:*
- 9: /SID22/gcgdata/geneseq/geneseq/NA1988.DAT:*
- 10: /SID22/gcgdata/geneseq/geneseq/NA1989.DAT:*
- 11: /SID22/gcgdata/geneseq/geneseq/NA1990.DAT:*
- 12: /SID22/gcgdata/geneseq/geneseq/NA1991.DAT:*
- 13: /SID22/gcgdata/geneseq/geneseq/NA1992.DAT:*
- 14: /SID22/gcgdata/geneseq/geneseq/NA1993.DAT:*
- 15: /SID22/gcgdata/geneseq/geneseq/NA1994.DAT:*
- 16: /SID22/gcgdata/geneseq/geneseq/NA1995.DAT:*
- 17: /SID22/gcgdata/geneseq/geneseq/NA1996.DAT:*
- 18: /SID22/gcgdata/geneseq/geneseq/NA1997.DAT:*
- 19: /SID22/gcgdata/geneseq/geneseq/NA1998.DAT:*
- 20: /SID22/gcgdata/geneseq/geneseq/NA1999.DAT:*
- 21: /SID22/gcgdata/geneseq/geneseq/NA2000.DAT:*
- 22: /SID22/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2301	100.0	12845	20 V84274	Mouse endothelial
2	55.4	2.4	5513	21 C63529	Slit protein codin
3	46	2.0	4064	21 A07587	Mouse piwi gene, d
4	45.6	2.0	51259	18 X83007	Partial mouse WRN
5	45	2.0	2362	21 Z50462	Human tissue trans
6	44	1.9	1110	19 V29343	Calcium ion channe
7	43.6	1.9	810	20 Z16495	Human gene express
8	43.2	1.9	11901	20 X02998	Human IL-1ra BAC c
9	42.6	1.9	8966	20 Z09581	Human Apo A1 genom
10	42.2	1.8	14507	18 T73568	Expression augment
11	42.2	1.8	14507	21 Z88869	Chinese hamster 2A

c 12	42	1.8	687	20	X37433	Human secreted pro
c 13	41.8	1.8	833	21	C98178	Human colon cancer
c 14	41.8	1.8	1751	19	V59610	Human secreted pro
c 15	41.8	1.8	41684	21	A28150	Human purH gene ge
c 16	41.6	1.8	10120	20	Z23683	Human DKC1 DNA fra
c 17	41.6	1.8	119950	20	X90201	Human yes1 gene.
c 18	41.4	1.8	32351	21	F21307	Human low adenosin
c 19	41.4	1.8	32351	21	A35185	Human adenosine re
c 20	41.4	1.8	40298	21	F21311	Human low adenosin
c 21	41.4	1.8	40298	21	A35189	Human adenosine re
c 22	41.2	1.8	1502	21	C59687	Human secreted pro
c 23	41	1.8	84607	20	X90847	Human PACAP genom
c 24	41	1.8	1082138	21	F22305	Arabidopsis thalia
c 25	40.8	1.8	837	20	X37525	Human secreted pro
c 26	40.8	1.8	1863	20	X85010	Human secreted pro
c 27	40.8	1.8	4116	20	X02983	Human IL-1ra BAC c
c 28	40.8	1.8	4736	20	X23315	Mouse l-alpha-Ohas
c 29	40.6	1.8	4034	20	X35011	Human glucose 6-ph
c 30	40.6	1.8	6511	16	Q95493	Human Cdn-2 DNA.
c 31	40.4	1.8	209	16	T26063	Human gene signatu
c 32	40.4	1.8	267	21	C29151	Human secreted pro
c 33	40.4	1.8	267	21	C29295	Human secreted pro
c 34	40.4	1.8	1838	20	X27332	Human secreted pro
c 35	40.4	1.8	4696	20	X03041	Human secreted pro
c 36	40.2	1.7	1035	21	C60001	Human secreted pro
c 37	40.2	1.7	6901	18	T78853	Human secreted pro
c 38	40.2	1.7	49999	20	Z23891	Human lecithin-cho
c 39	40.2	1.7	49999	20	Z23896	Murine LOBO genom
c 40	40.2	1.7	138169	21	A34791	Murine LOBO homolo
c 41	40.2	1.7	141589	21	F20913	Human adenosine re
c 42	40.2	1.7	141589	21	F21127	Human ELAM-1 polyn
c 43	40.2	1.7	141589	21	F21152	Human low adenosin
c 44	40.2	1.7	141589	21	A35005	Human low adenosin
c 45	40.2	1.7	141589	21	A35030	Human adenosine re

ALIGNMENTS

RESULT 1
V84274
ID V84274 standard; DNA; 12845 BP.
XX V84274;
AC
XX
XX
DF 12-APR-1999 (first entry)
XX
DE Mouse endothelial growth factor receptor-2 gene Flk-1 5' region.
XX
KW Flk-1; endothelial growth factor receptor-2; VEGF; transcription;
KW endothelium; enhancer; vascular disease; angiogenesis; cancer;
KW diabetic retinopathy; rheumatoid arthritis; wound healing;
KW vulnery; atherosclerosis; tumour; neuronal disorder; therapy;
KW diagnosis; mouse; ss.
XX Mus sp.
XX
XX
XX Key Location/Qualifiers
FT promoter 6036..6959
FT exon /*tag= a
6661..7026
FT /*tag= b
FT /*number= 1
FT /*codon_start= 6960
FT intron 7027..10632
FT /*tag= c
FT /*number= 1
FT /*note= "DNA sequences comprising nucleotides
8260-10560, 8336-10608 and/or 10094-10608
are specifically claimed in Claim 3b"
FT protein_bind 10137..10155
FT /*tag= d
FT /*note= "GATA/PEA3 consensus binding site"

FT protein_bind 10166..10179
 FT /*tag= e
 FT /note= "GATA consensus binding site"
 FT 10187..10197
 FT /*tag= f
 FT /note= "AP1 consensus binding site"
 FT 10252..10261
 FT /*tag= g
 FT /note= "NFKB consensus binding site"
 FT 10273..10283
 FT /*tag= h
 FT /note= "AP1 consensus binding site"
 FT 10301..10309
 FT /*tag= i
 FT /note= "SPAT consensus binding site"
 FT 10356..10372
 FT /*tag= j
 FT /note= "Ets-1/GATA consensus binding site"
 FT 10395..10404
 FT /*tag= k
 FT /note= "Ets-1 consensus binding site"
 FT 10453..10477
 FT /*tag= l
 FT /note= "SCL/TAL-1 Ets-1 consensus binding site"
 FT 10843..10726
 FT /*tag= m
 FT /number= 2
 FT 10727..12673
 FT /*tag= n
 FT /number= 2
 FT 12674..12845
 FT /*tag= o
 FT /number= 3
 FT
 XX WO9855638-A1.
 XX
 XX
 PD 10-DEC-1998.
 XX
 PF 03-JUN-1998; 98WO-EP03318.
 XX
 PR 03-JUN-1997; 97EP-0108959.
 XX
 XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 XX
 PI Breier G, Risau W, Roenicke V;
 XX
 DR WPI; 1999-059915/05.
 XX
 XX New recombinant DNA containing heterologous sequence controlled by
 PT endothelial cell specific regulator - from the Flk-1 gene, used to
 PT treat, prevent or diagnose vascular disease, tumours, also to screen
 PT for transcriptional regulators
 XX
 XX Claim 3; Fig 1; 107pp; English.
 XX
 CC This 12.8 kb DNA sequence spans the region from -6.65 kb relative to
 CC the transcriptional start site to +6.15 (located in the third exon)
 CC of the murine endothelial growth factor (VEGF) receptor-2 gene
 CC Flk-1. This portion of the Flk-1 gene was isolated from the liver
 CC of mouse strain 129/SvJ by screening a phage library of liver
 CC genomic DNA in vector lambda FixII. Sequences within the 5'-flanking
 CC region of the gene, in combination with sequences located within the
 CC first intron, specifically and reproducibly target expression of
 CC heterologous DNA to angioblasts during early stages of vascular
 CC development and also to the vasculature of postnatal mice. The
 CC regulatory sequences of the first intron also function as an
 CC autonomous endothelium-specific enhancer when fused to a
 CC heterologous promoter (e.g. HSV-TK promoter). This Flk-1 intron
 CC enhancer contains several potential binding sites for transcription
 CC factors of the Ets and GATA families. The invention provides a
 CC new claimed recombinant DNA (I) comprising at least one regulatory
 CC sequence from an intron of the Flk-1 gene, or its homologue, to
 CC control expression in endothelial cells, in vivo, linked to (b) a

CC heterologous DNA. (I) is used to produce transgenic animals, and
 CC these, or transformed cells, are used to identify agents (A),
 CC potential pharmaceuticals, that suppress, activate or enhance
 CC transcription of genes in endothelial cells. (I), vectors and (A)
 CC are used to direct or prevent (for antisense sequences) expression
 CC of genes specifically in endothelial cells, e.g. for treating
 CC angiogenesis, cancer, diabetic retinopathy, rheumatoid arthritis
 CC etc., and in wound healing, particularly for treating vascular
 CC disease (particularly proliferation of smooth muscle cells,
 CC specifically atherosclerosis), tumours and neuronal disorders.
 CC They may also be used to induce vascular disease in the transgenic
 CC animals or diagnostically, particularly for studying (mal)function,
 CC interactions and unregulated expression of endothelial products.
 CC (I) provide specific modulation of gene expression in endothelial
 CC cells, at all stages of development.
 XX
 SQ Sequence 12845 BP; 3321 A; 2893 C; 2863 G; 3723 T; 45 other;

Query Match 100.0%; Score 2301; DB 20; Length 12845;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2301; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgaataagatggaggtgcccctttggaggtgtgtgacttcttaattggattgggctata 60
 DB 8260 tgaataagatggaggtgcccctttggaggtgtgtgacttcttaattggattgggctata 8319
 QY 61 attgtgccatccaaagtctcgagacagagccgtgtgtttcttctctgtcttgaagc 120
 DB 8320 attgtgccatccaaagtctcgagacagagccgtgtgtttcttctctgtcttgaagc 8379
 QY 121 gggaaggataacagtgacacaaataatattgtgtgttatcgatttgacataaaaggg 180
 DB 8380 gggaaggataacagtgacacaaataatattgtgtgttatcgatttgacataaaaggg 8439
 QY 181 cttttattgtatagatagatatatctcttgcagtcagaatagctgtctataaagacag 240
 DB 8440 cttttattgtatagatagatatatctcttgcagtcagaatagctgtctataaagacag 8499
 QY 241 aacccaaactgccgatgaaatgaatgaggttttaataaaggcgatggatgagcattagt 300
 DB 8500 aacccaaactgccgatgaaatgaatgaggttttaataaaggcgatggatgagcattagt 8559
 QY 301 cactgatgaaatctccagttattgataacctcattgactgactgattgattgacagactgt 360
 DB 8560 cactgatgaaatctccagttattgataacctcattgactgactgattgattgacagactgt 8619
 QY 361 attgtatgggcatccttttaagatgagcatagcacaacgtgcctgcactcaagaagaat 420
 DB 8620 attgtatgggcatccttttaagatgagcatagcacaacgtgcctgcactcaagaagaat 8679
 QY 421 ctatggctgtatgtattacagagacagttgagaagctcttagtggctctggtgtgtaga 480
 DB 8680 ctatggctgtatgtattacagagacagttgagaagctcttagtggctctggtgtgtaga 8739
 QY 481 tcagcggtagagcgtgaggtctgcgtcgtctctcctgcactgagaataaaagccatt 540
 DB 8740 tcagcggtagagcgtgaggtctgcgtcgtctctcctgcactgagaataaaagccatt 8799
 QY 541 tactgtggtgtgagtgaggtggcgagttgtgacgagttactactacatttctccacaca 600
 DB 8800 tactgtggtgtgagtgaggtggcgagttgtgacgagttactactacatttctccacaca 8859
 QY 601 tctgcctgactaatgagttcatcagatgagcgttatccagtgattgttttcaggttaattgg 660
 DB 8860 tctgcctgactaatgagttcatcagatgagcgttatccagtgattgttttcaggttaattgg 8919
 QY 661 tctcagtcagtgttagaattctactatcaacaataattgtttctcatttctctctt 720
 DB 8920 tctcagtcagtgttagaattctactatcaacaataattgtttctcatttctctctt 8979
 QY 721 ctcaacaagaatgaagttccattatgaaagcgtttgtaaagacatttaacgtcttcg 780

Db 8980 ctcaacaaagtaagattccattattgaagagctgtgtttaagagacatttttaactgtctgc 9039
 Qy 781 ctatgttaggacagtgacttatttcatattgacaaatattatgccgattaatgaatat 840
 Db 9040 ctatgttaggacagtgacttatttcatattgacaaatattatgccgattaatgaatat 9099
 Qy 841 gactaccagttctatagctgtctcaaggcagacaaagagacatctgtgatccagtcactt 900
 Db 9100 gactaccagttctatagctgtctcaaggcagacaaagagacatctgtgatccagtcactt 9159
 Qy 901 taaatgccatttaaaatgcaataatttgttggtctaggaaataaacacacactgtaaaagttag 960
 Db 9160 taaatgccatttaaaatgcaataatttgttggtctaggaaataaacacacactgtaaaagttag 9219
 Qy 961 aatcacggcccaaacacaaagcttttaacaatgccaaactagcttctgagattcattaatgt 1020
 Db 9220 aatcacggcccaaacacaaagcttttaacaatgccaaactagcttctgagattcattaatgt 9279
 Qy 1021 catttaattaccaatgttttaaaatattgtcttaataatgcttaataatctatagttgaacag 1080
 Db 9280 catttaattaccaatgttttaaaatattgtcttaataatgcttaataatctatagttgaacag 9339
 Qy 1081 caacacatgtacattcttataaagttgggtattattcagggtggcagtagctgtagactattg 1140
 Db 9340 caacacatgtacattcttataaagttgggtattattcagggtggcagtagctgtagactattg 9399
 Qy 1141 cacatctgtgtgtgagccagtgagaaactgctctctgctgttctcagaagagccacag 1200
 Db 9400 cacatctgtgtgtgagccagtgagaaactgctctctgctgttctcagaagagccacag 9459
 Qy 1201 tgtcacggcattggctatttgccttgctgtcttcttgaataactttattgacatggcctaat 1260
 Db 9460 tgtcacggcattggctatttgccttgctgtcttcttgaataactttattgacatggcctaat 9519
 Qy 1261 ctctgttcaagttcacttatttgcacacaaactgcaatgcagctgagcccttaaggagtc 1320
 Db 9520 ctctgttcaagttcacttatttgcacacaaactgcaatgcagctgagcccttaaggagtc 9579
 Qy 1321 atctgttcttagtcagtcgaatagaaagcctggatgctgtgctgtctgtatttaattagttta 1380
 Db 9580 atctgttcttagtcagtcgaatagaaagcctggatgctgtgctgtctgtatttaattagttta 9639
 Qy 1381 ttcttcttcttgagacagagtcactgtgtgtgcccaggtagcttcaaaacttgcggctc 1440
 Db 9640 ttcttcttcttgagacagagtcactgtgtgtgcccaggtagcttcaaaacttgcggctc 9699
 Qy 1441 catttcttcaactcatcagaatgctggcttccagggtgtgtgcccacactagtagctc 1500
 Db 9700 catttcttcaactcatcagaatgctggcttccagggtgtgtgcccacactagtagctc 9759
 Qy 1501 gcgttttaagctaaagagctggaagatcctgatgtccctttaccatgggtggcagattgtacag 1560
 Db 9760 gcgttttaagctaaagagctggaagatcctgatgtccctttaccatgggtggcagattgtacag 9819
 Qy 1561 gttagtgtactgaaactagttatctcgtgtgtgttaatgacotgagctgggtatgtatctct 1620
 Db 9820 gttagtgtactgaaactagttatctcgtgtgtgttaatgacotgagctgggtatgtatctct 9879
 Qy 1621 caaagatcttttttgcatttcaactcagttaggttaacaaagttcttlaagctccacagcttgg 1680
 Db 9880 caaagatcttttttgcatttcaactcagttaggttaacaaagttcttlaagctccacagcttgg 9939
 Qy 1681 attggcagctagcagagcttggatttaagttgagttgggaccccttagctattgtcattag 1740
 Db 9940 attggcagctagcagagcttggatttaagttgagttgggaccccttagctattgtcattag 9999
 Qy 1741 acttacactatttttagtttgcctgtgagtttatgaatatgcatgtatgcatgaacttgg 1800
 Db 10000 acttacactatttttagtttgcctgtgagtttatgaatatgcatgtatgcatgaacttgg 10059
 Qy 1801 gagatatttttctcccccaattctcttctccatttaataatgtctgtctttagaagcca 1860
 Db 10060 gagatatttttctcccccaattctcttctccatttaataatgtctgtctttagaagcca 10119

Qy 1861 ctgctcagcttctgagctcagatcaccaaaaggaagtctgtgtacacagcatgataaaaga 1920
 Db 10120 ctgctcagcttctgagctcagatcaccaaaaggaagtctgtgtacacagcatgataaaaga 10179
 Qy 1921 caatgggacggggtcacagtgagctcccttcccttccaggggtatggagacgagctgtaga 1980
 Db 10180 caatgggacggggtcacagtgagctcccttcccttccaggggtatggagacgagctgtaga 10239
 Qy 1981 gagatgtctccagggaggttttccatttaacagcaatttagtcagatctgtgcatcctatgc 2040
 Db 10240 gagatgtctccagggaggttttccatttaacagcaatttagtcagatctgtgcatcctatgc 10299
 Qy 2041 tttaacagaataatgtcagtgggcctgagatcatcagatgaggttcatcggggttttcaatgt 2100
 Db 10300 tttaacagaataatgtcagtgggcctgagatcatcagatgaggttcatcggggttttcaatgt 10359
 Qy 2101 cccgtatccttttgaagaccttgaagtgtgcaacgcgggaaaaacaggaactccacctg 2160
 Db 10360 cccgtatccttttgaagaccttgaagtgtgcaacgcgggaaaaacaggaactccacctg 10419
 Qy 2161 gtgcggtgaattgcagagctgt 2220
 Db 10420 gtgcggtgaattgcagagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 10479
 Qy 2221 tgacagagcttgtgaacttttaactgggactgggcaaggtcaaatccacacctttatacaat 2280
 Db 10480 tgacagagcttgtgaacttttaactgggactgggcaaggtcaaatccacacctttatacaat 10539
 Qy 2281 gaattgctgaagagggcctttt 2301
 Db 10540 gaattgctgaagagggcctttt 10560

RESULT 2

C63529/c
 ID C63529 standard; DNA; 5513 BP.

XX C63529;

XX 08-FEB-2001 (first entry)

XX Slit protein coding sequence.

DE Slit protein; frog; cell migration; neuronal; nerve axon; dendrite;
 KW leukocyte; graft rejection; HIV infection; antinflammatory;
 KW wound repair; organ regeneration; asthma; arthritis; glomerulonephritis;
 KW cystic fibrosis; ulcerative colitis; Crohn's disease; multiple sclerosis;
 KW allergic encephalomyelitis; Alzheimer's disease;
 KW coronary artery restenosis; ss.

OS Xenopus sp.

PN WO200055321-A2.

XX 21-SEP-2000.

PF 16-MAR-2000; 2000WO-US07040.

XX 17-MAR-1999; 99US-0124767.

PA (UNIW) UNIV WASHINGTON.

PI Rao Y, Wu JY;

DR WPI; 2000-628199/60.

XX P-PSDB; B28151.

PT New vertebrate DNA encoding a slit protein that is useful for treating
 PT graft rejection, inhibiting infection of a cell by human
 PT immunodeficiency virus, inhibiting inflammation, and for aiding wound
 PT repair and organ regeneration -
 XX

CC The products can be used for the detection and treatment of Werner's
 CC Syndrome (WS), an autosomal recessive disorder with a complex phenotype,
 CC as well as related diseases.

XX Sequence 51259 BP; 14533 A; 9635 C; 10266 G; 16825 T; 0 other;

Query Match 2.0%; Score 45.6; DB 18; Length 51259;
 Best Local Similarity 62.1%; Pred. No. 0.067;
 Matches 72; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 1368 tattaattagttattcttctctgagacagatctcaactgtgtgcccaggctagtctct 1427

Db 23743 tttagtttagtttactttttttgacagggctcactgtgtagctggggacaagct 23802

Qy 1428 caaactgcggctcattgtctcaactcatcagaatgctgggctccagggtgtgtgc 1483

Db 23803 ccaccctgtccctcttctcaacctctgagctgctgggatacacaggcgtgtgc 23858

RESULT 5

ID Z50462/c
 ID Z50462 standard; mRNA; 2362 BP.

AC Z50462;

XX 23-MAY-2000 (first entry)

XX Human tissue transglutaminase (tTGase) homologue-3 encoding mRNA.

XX Tissue transglutaminase; tTGase; human; homologue; myeloid cell; LTR-HPC;
 KW long-term repopulating hematopoietic progenitor cell; calcium dependent;
 KW apoptosis; cell adhesion; metastasis; extracellular matrix assembly;
 KW ECM; marker; allogeneic transplant; gene therapy; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
 FH CDS 133..1182

FT /*tag= a

FT /product= "Human tissue transglutaminase homologue-3"

FT /note= "GTP binding calcium dependent enzyme"

XX W0200006766-A1.

XX 10-FEB-2000.

XX 28-JUL-1999; 99WO-US17096.

XX 29-JUL-1998; 98US-0123970.

XX (NYBL-) NEW YORK BLOOD CENT INC.

XX Visser JWM, Ivanova NB;

XX WPI; 2000-195317/17.

XX P-PSDB; Y44919.

XX Differentially identifying long-term repopulating hematopoietic
 PT progenitor cells using tissue transglutaminase, useful for improving
 PT medical procedures, including gene therapy

XX Disclosure; Page 48-50; 59pp; English.

XX The patent discloses a method of differentially identifying long-term
 CC repopulating hematopoietic progenitor cells (LTR-HPC), by comparing the
 CC expression of tissue transglutaminase (tTGase), a calcium dependent
 CC enzyme, in myeloid cells like blood or marrow to a standard. A higher
 CC level of tTGase expression compared to background expression, indicates
 CC the presence of LTR-HPC in the sample. Tissue TGase catalyses post-
 CC translational modification of proteins and is involved in apoptosis,
 CC cell adhesion, metastasis and extracellular matrix (ECM) assembly. It
 CC acts as a marker that distinguishes the LTR-HPC from other (im)mature

CC blood cells and is also useful for improved quantification of a mammal's
 CC hematopoietic competence. The method is useful for functional selection
 CC of progenitor cells, enabling allogeneic transplant procedures and ex
 CC vivo manipulation of HPC for use in gene therapy. The present sequence
 CC is an mRNA encoding a homologue of human tTGase, expressed in LTR-HPC.

XX Sequence 2362 BP; 616 A; 593 C; 640 G; 513 T; 0 other;

Query Match 2.0%; Score 45; DB 21; Length 2362;
 Best Local Similarity 70.6%; Pred. No. 0.019;
 Matches 60; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy 1356 atgcctgcctctattatagttattctctctctctgagacagatctcactgtgtgc 1415

Db 1787 ATGCTGCTAAATTTTATTTATTTTATTTTGTAGACAGAGCTCTACTATATGCG 1728

Qy 1416 ccaggctagctcaactgcggctc 1440

Db 1727 CCAGGCTGGTCTCAAACTCCGGCC 1703

RESULT 6

V29343/c

ID V29343 standard; DNA; 1110 BP.

XX AC V29343;

XX 31-JUL-1998 (first entry)

XX Calcium ion channel alpha subunit exon 15/intron partial sequence.

XX Calcium ion channel alpha subunit; human; episodic ataxia type 2;

KW familial hemiplegic migraine; FHM; EA-2; treatment; diagnosis;

KW exon; intron; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
 FH intron 1..201

FT /*tag= a

FT /number= 14

FT /note= "partial sequence"

FT 202..274

FT /*tag= b

FT /number= 15

FT 275..1110

FT /*tag= c

FT /number= 15

FT /note= "partial sequence"

XX EP834561-A1.

XX 08-APR-1998.

XX 27-SEP-1996; 96EP-0202707.

XX 27-SEP-1996; 96EP-0202707.

XX (UYLE-) RIJKSUNIV LEIDEN.

XX Ferrari MD, Frants RRIE, Ophoff RA, Terwindt GM;

XX WPI; 1998-195461/18.

XX New human nucleic acid associated with migraine and episodic ataxia
 PT type 2 - useful for diagnosis and development of specific treatments

XX Disclosure; Fig 1; 157pp; English.

XX Sequences shown in V29330 to V29371 represent the 47 exons and flanking
 CC intronic sequences containing the complete coding region of the human
 CC calcium ion channel alpha 1 subunit gene and part of untranslated


```

FT misc_feature 10100..14923
FT /tag= g
FT /note= "EASE (Claim 1)"
FT misc_feature 14290..14507
FT /tag= h
FT /note= "the EASE sequence is preferably ligated
FT to DNA comprising nucleotides 14290-14507
FT (Claim 3)"
XX
XX WO9725420-A1.
XX
XX
XX
XX
XX 17-JUL-1997.
XX
XX 10-JAN-1997; 97WO-US00483.
XX
XX 11-JAN-1996; 96US-0586509.
XX
XX (IMMV ) IMMUNEX CORP.
XX
XX PI Lee C, Morris AE, Thomas JN;
XX WPI; 1997-372861/34.
XX
XX
XX Expression augmenting sequence elements - used in vectors for
XX expressing recombinant proteins at high levels in shorter periods of
XX time
XX
XX Claim 1; Page 23-30; 36pp; English.
XX
XX A 14507 bp DNA sequence (773568) comprises a novel transcription
XX regulatory element, expression augmenting sequence element (EASE),
XX that facilitates high expression of recombinant proteins in
XX mammalian host cells. It was identified by cloning the integration
XX site of a unique expression cassette encoding recombinant dimeric
XX tumour necrosis factor receptor-immunoglobulin Fc fusion protein
XX from genomic DNA of a CHO 2A5-3 cell line expressing this protein
XX at a high level. Expression vectors incorporating an EASE show
XX a 2- to 8-fold improvement of recombinant protein expression levels
XX in CHO host cells. The EASE sequence also facilitates high
XX expression in shorter periods of time.
XX
XX Sequence 14507 BP; 4225 A; 2548 C; 3136 G; 4598 T; 0 other;

Query Match 1.8%; Score 42.2; DB 18; Length 14507;
Best Local Similarity 59.9%; Pred. No. 0.31;
Matches 88; Conservative 0; Mismatches 58; Indels 1; Gaps 1;

QY 1364 ctgctataattagttattcttctcttgagacagagctcactgtgtgcccaggcta 1423
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
146 cagcgaagattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtg 205

QY 1424 gtctcaact-tgcggtccattgtctcactcactcagagctcaggtgtgtgtg 1482
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
206 tccaggactcagagatccactcctctgtctctgtgtgtgtgtgtgtgtgtgtgtg 265

QY 1483 caccacactaggtagctgcggttttaa 1509
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
266 ctgactacaggaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtg 292

RESULT 11
Z88869
ID Z88869 standard; DNA; 14507 BP.
XX
AC Z88869;
XX
XX Chinese hamster 2A5-3 lambda EASE DNA.
XX
XX EASE; expression augmenting sequence element; chinese hamster; ds.
XX

```

```

OS Cricetus sp.
XX
XX US6027915-A.
XX
XX 22-FEB-2000.
XX
XX 13-JAN-1997; 97US-0785150.
XX
XX 11-JAN-1996; 96US-0586509.
XX
XX (IMMV ) IMMUNEX CORP.
XX
XX Morris AE, Thomas JN, Lee C;
XX WPI; 2000-194831/17.
XX
XX Improved expression of recombinant proteins in stable cell pools of
XX mammalian cells comprises transforming cells with a vector containing
XX an expression augmenting sequence element -
XX
XX Claim 1; Column 17-30; 18pp; English.
XX
XX This invention describes a novel expression augmenting sequence element
XX (EASE) (I) isolated from Chinese hamster ovary cells. (I) facilitates
XX high expression of recombinant proteins in mammalian host cells in a
XX short time period. (I) is used for improving expression of recombinant
XX proteins 2-8 fold in stable cell pools when it is present in an
XX expression vector. (I) improves recombinant protein expression in
XX mammalian cells and facilitates the use of mammalian cells in
XX recombinant protein production.
XX
XX Sequence 14507 BP; 4225 A; 2548 C; 3136 G; 4598 T; 0 other;

Query Match 1.8%; Score 42.2; DB 21; Length 14507;
Best Local Similarity 59.9%; Pred. No. 0.31;
Matches 88; Conservative 0; Mismatches 58; Indels 1; Gaps 1;

QY 1364 ctgctataattagttattcttctcttgagacagagctcactgtgtgcccaggcta 1423
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
146 cagcgaagattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtg 205

QY 1424 gtctcaact-tgcggtccattgtctcactcactcagagctcaggtgtgtgtg 1482
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
206 tccaggactcagagatccactcctctgtctctgtgtgtgtgtgtgtgtgtgtgtgtg 265

QY 1483 caccacactaggtagctgcggttttaa 1509
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
266 ctgactacaggaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtg 292

RESULT 12
X37433/C
ID X37433 standard; cDNA; 687 BP.
XX
AC X37433;
XX
XX 06-JUL-1999 (first entry)
XX
XX Human secreted protein cDNA fragment containing gene 65.
XX
XX Human; secreted protein; prevention; treatment; protein therapy;
XX gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
XX developmental abnormality; foetal deficiency; blood disorder; lymphoma;
XX leukemia; immune system disorder; autoimmune disease; hepatic disease;
XX renal disease; inflammation; allergy; asthma; sepsis; diabetes; AIDS;
XX Alzheimer's disease; cognitive disorder; schizophrenia; osteoporosis;
XX arthritis; psoriasis; digestive; endocrine; infection; ss.
XX
XX Homo sapiens.
XX
XX WO9909155-A1.
XX

```


XX 06-JAN-1999 (first entry)
DT Human secreted protein gene 100 clone HLQAB52.
XX
DE
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
XX WO9839448-A2.
XX
XX 11-SEP-1998.
XX
XX 06-MAR-1998; 98WO-US04493.
XX
XX 02-OCT-1997; 97US-0061060.
PR 23-MAY-1997; 97US-0047615.
PR 23-MAY-1997; 97US-0047617.
PR 23-MAY-1997; 97US-0047618.
PR 23-MAY-1997; 97US-0047632.
PR 23-MAY-1997; 97US-0047633.
PR 06-JUN-1997; 97US-0048964.
PR 06-JUN-1997; 97US-0048974.
PR 13-JUN-1997; 97US-0049610.
PR 08-JUL-1997; 97US-0051926.
PR 16-JUL-1997; 97US-0052874.
PR 18-AUG-1997; 97US-0055724.
PR 22-AUG-1997; 97US-0056630.
PR 22-AUG-1997; 97US-0056631.
PR 22-AUG-1997; 97US-0056632.
PR 22-AUG-1997; 97US-0056636.
PR 22-AUG-1997; 97US-0056637.
PR 22-AUG-1997; 97US-0056662.
PR 22-AUG-1997; 97US-0056664.
PR 22-AUG-1997; 97US-0056845.
PR 22-AUG-1997; 97US-0056862.
PR 22-AUG-1997; 97US-0056864.
PR 22-AUG-1997; 97US-0056872.
PR 22-AUG-1997; 97US-0056874.
PR 22-AUG-1997; 97US-0056875.
PR 22-AUG-1997; 97US-0056876.
PR 22-AUG-1997; 97US-0056877.
PR 22-AUG-1997; 97US-0056878.
PR 22-AUG-1997; 97US-0056879.
PR 22-AUG-1997; 97US-0056880.
PR 22-AUG-1997; 97US-0056881.
PR 22-AUG-1997; 97US-0056882.
PR 22-AUG-1997; 97US-0056884.
PR 22-AUG-1997; 97US-0056886.
PR 22-AUG-1997; 97US-0056887.
PR 22-AUG-1997; 97US-0056888.
PR 22-AUG-1997; 97US-0056889.
PR 22-AUG-1997; 97US-0056892.
PR 22-AUG-1997; 97US-0056893.
PR 22-AUG-1997; 97US-0056894.
PR 22-AUG-1997; 97US-0056903.
PR 22-AUG-1997; 97US-0056908.
PR 22-AUG-1997; 97US-0056909.
PR 22-AUG-1997; 97US-0056910.
PR 22-AUG-1997; 97US-0056911.
PR 05-SEP-1997; 97US-0057650.
PR 05-SEP-1997; 97US-0057669.
PR 05-SEP-1997; 97US-0057761.
PR 12-SEP-1997; 97US-0058785.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Bedharik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;
PI Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS;
PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
XX
XX WPI; 1998-506364/43.
DR P-PSDB; W74830.
XX
XX New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX
XX Claim 1; Page 333-334; 721pp; English.
PS
XX This sequence represents a nucleic acid molecule designated Gene 101 from
CC the human cDNA clone HLQAB52 (deposited as clone ATCC 97901 and ATCC
CC 209047) which encodes a secreted human protein. The gene can be used to
CC generate fusion proteins by linking to the gene to a human immunoglobulin
CC Fc portion (e.g. V59502) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 186 novel genes and their fragments (nucleic
CC acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 13:30:07 ; Search time 150.46 Seconds
(without alignments)
2833.205 Million cell updates/sec

Title: US-09-445-201-1-copy_8260_10560

Perfect score: 2301
Sequence: 1 tgaataagatgaggtgcc.....aattgctgaagagcctttt 2301

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PTCUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47.6	2.1	7218	1	US-08-232-463-14
2	45.6	2.0	51259	3	US-08-781-891-209
3	43.2	1.9	152331	4	US-09-128-155-16
4	42.2	1.8	14507	3	US-08-785-150-1
5	40.8	1.8	152331	4	US-09-128-155-16
6	40.4	1.8	176373	4	US-09-128-155-16
7	40.2	1.7	380	1	US-08-126-587C-5
8	40	1.7	1920	1	US-08-087-772A-1
9	39.6	1.7	246240	2	US-08-724-394A-20
10	39.6	1.7	246240	2	US-08-724-394A-21
11	39.6	1.7	246240	2	US-08-724-394A-22
12	39.4	1.7	2880	4	US-09-115-954-3
13	39.4	1.7	3842	4	US-09-115-954-7
14	39.4	1.7	3912	4	US-09-115-954-1
15	39.2	1.7	11725	2	US-08-756-506-1
16	39.2	1.7	17041	1	US-08-076-011-1
17	39	1.7	35060	3	US-08-814-095-7
18	38.6	1.7	2908	3	US-08-487-799-1
19	38.6	1.7	6769	1	US-08-480-784-20
20	38.6	1.7	6769	1	US-08-483-553-20
21	38.6	1.7	6769	1	US-08-487-002-20
22	38.6	1.7	6769	1	US-08-483-554B-20
23	38.6	1.7	6769	1	US-08-488-011B-20
24	38.6	1.7	6769	4	US-08-850-727-20
25	38.6	1.7	6769	5	PCT-US95-10202-20
26	38.6	1.7	6769	5	PCT-US95-10203-20
27	38.6	1.7	6769	5	PCT-US95-10220-20

c	28	38.4	1.7	501	3	US-08-699-628-1	Sequence 1, Appli
	29	38.4	1.7	2520	2	US-08-454-557C-50	Sequence 50, Appl
	30	38.4	1.7	2520	2	US-08-340-426D-50	Sequence 50, Appl
	31	38.4	1.7	2520	2	US-08-450-673C-50	Sequence 50, Appl
	32	38.4	1.7	2520	5	PCT-US95-17111A-50	Sequence 50, Appl
c	33	38.2	1.7	5095	1	US-08-092-817-3	Sequence 3, Appli
	34	38	1.7	31571	1	US-08-323-443B-1	Sequence 1, Appli
c	35	38	1.7	40352	3	US-08-846-111D-15	Sequence 15, Appl
	36	38	1.7	53526	3	US-08-658-136-2	Sequence 1, Appli
	37	38	1.7	53577	3	US-08-658-136-1	Sequence 1, Appli
	38	37.6	1.6	615	4	US-08-998-416-186	Sequence 186, App
	39	37.4	1.6	7720	4	US-09-318-448-5	Sequence 5, Appli
	40	37	1.6	4084	3	US-08-866-340-1	Sequence 1, Appli
	41	37	1.6	4460	4	US-09-103-875-4	Sequence 4, Appli
	42	36.8	1.6	4922	2	US-08-330-272-5	Sequence 5, Appli
	43	36.8	1.6	4922	5	PCT-US95-13663-5	Sequence 5, Appli
c	44	36.6	1.6	5375	4	US-08-757-223-7	Sequence 7, Appli
c	45	36.6	1.6	10825	3	US-08-652-265-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pT29pt-F1s
US-08-232-463-14


```

Query Match      1.8%; Score 42.2; DB 3; Length 14507;
Best Local Similarity 59.9%; Pred. No. 0.043;
Matches 88; Conservative 0; Mismatches 58; Indels 1; Gaps 1;

QY 1364 ctgctattaattagtaattctctctctgagacagagtctcactgtgtggtgcccaggcta 1423
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 146 CAGCCACAGATTGTGTGTTGTTGTTCCGAGAAAGGGTTCTCTGTGTAGCCCTTCGCTG 205
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1424 gtctaaact-tgcggtccatttgtctcactcatcagaatgcctggtgtctccaggtgtgtg 1482
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 206 TCAGGAACTCAGAGATCCACCTGCCCTCTGCTTCCCTGAGTGTGGGATTAAGAGTGTGTG 265
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1483 caccacactaggtagctcgcggttttaa 1509
      | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 266 CTGACTACAGCAAGCTTGTTGTTTA 292
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
US-09-128-155-16/c
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02

```

	Query Match	1.8%	Score 40.8	DB 4	Length 152331
	Best Local Similarity	61.1%	Pred. No. 0.47		
	Matches 66	Conservative 0	Mismatches 42	Indels 0	Gaps 0
QY 1326	ttcttagtcagtgccgaattagaagccctgagtcgctgcctgctattaattagttattcttt				
Db 54518	TTCTTGAGTAGGTAGGATTACAGGCACGCACCGCCTCTGCGCTTTTTTTTTTTTTTTT				
QY 1386	ctctctcgaacgagtgctcactgctgtggcccgagctagtcctcaact				
Db 54458	TTTTTTAGAGATGGGGTCTCACTATGTTGCCAGGCCAGTTTTTCAAAAT				

```

Query Match      1.8%; Score 40.4; DB 4; Length 176373;
Best Local Similarity 59.6%; Pred. No. 0.68;
Matches 68; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 1378 ttattctctcttgagacagagtcctcaactgtgtgcccaggctagtctcaaaactgcg 1437
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61378 TTTTCTTTTCTTTTGTGACACGAGATCTCGCTCTCTGCGCCACAGCGGGACTCGGAGCTGCA 61319

Qy 1438 gtccatttgtctcactcatcagaatgctgggctttccagggtgtgtgcaaccact 1491
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61318 GTGGCGCAATCTCGGCTCACTGCAAGTCCGCTTCCCGGGTTTCACGGCACTCT 61265

RESULT 7
US-08-126-587C-5/c
; Sequence 5, Application US/08126587C
; Patent No. 5534438

```

RESULT 8
US-08-087-772A-1/c
; Sequence 1, Application US/08087772A
; Patent No. 5691155
; GENERAL INFORMATION:
; APPLICANT: Nahmias, Clara
; APPLICANT: Emorine, Jean L.
; APPLICANT: Strosberg, Donny A.
; TITLE OF INVENTION: Nucleotide Sequences Encoding the Murine
; TITLE OF INVENTION: Beta3-Adrenergic Receptor and Their Applications
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Versi
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/724,394A
 FILING DATE: 01-OCT-1996
 CLASSIFICATION: 536
 ATTORNEY/AGENT INFORMATION:
 NAME: Fitts, Renee A.
 REGISTRATION NUMBER: 35 136
 REFERENCE/DOCKET NUMBER: 017957-000100
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-576-0200
 TELEFAX: 415-576-0300
 INFORMATION FOR SEQ ID NO: 22:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 246240 base pairs

```
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-22

Query Match
Best Local Similarity 1.7%; Score 39.6; DB 2; Length 246240;
Matches 54; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 1363 cctgtattattagttattcttctcttgagacagagtcactgtgtgcccaggtc 1422
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 95222 CCATTTTCTTTTCTTCTCTCTTTTGTAGACAGGGCTCACTATGTTGCCAGGCT 95281

QY 1423 agtctcaaaacttgcggtc 1440
||||| ||||| |||
Db 95282 GGTCTCCAACTTCTGGGC 95299

RESULT 12
US-09-115-954-3/c
; Sequence 3, Application US/09115954B
; Patent No. 6200776
; GENERAL INFORMATION:
; APPLICANT: Boron, Walter F
; APPLICANT: Brill, Antoine M
; APPLICANT: Khandoudi, Nassirah
; APPLICANT: Martin, Xavier
; APPLICANT: Jupe, Steven C
; APPLICANT: Rawlings, Christopher J
; APPLICANT: Doe, Trudy R
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GH-30409
; CURRENT APPLICATION NUMBER: US/09/115,954B
; CURRENT FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: EP97401714.7
; EARLIER FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2880
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-115-954-3

Query Match
Best Local Similarity 1.7%; Score 39.4; DB 4; Length 2880;
Matches 52; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1368 tattaattagttattcttctcttgagacagagtcactgtgtgcccaggtc 1427
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2853 TTTTCTTTTCTTTTCTTTTCTTTTGTAGACAGGGCTCTCATTTATGCTAGGCTAGTCT 2794

QY 1428 caaacttgcggtc 1440
||||| ||||| |||
Db 2793 CAAACTTCTGGCC 2781

RESULT 13
US-09-115-954-7/c
; Sequence 7, Application US/09115954B
; Patent No. 6200776
; GENERAL INFORMATION:
; APPLICANT: Boron, Walter F
; APPLICANT: Brill, Antoine M
; APPLICANT: Khandoudi, Nassirah
; APPLICANT: Martin, Xavier
```

```
; APPLICANT: Jupe, Steven C
; APPLICANT: Rawlings, Christopher J
; APPLICANT: Doe, Trudy R
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GH-30409
; CURRENT APPLICATION NUMBER: US/09/115,954B
; CURRENT FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: EP97401714.7
; EARLIER FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 3842
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-115-954-7

Query Match
Best Local Similarity 1.7%; Score 39.4; DB 4; Length 3842;
Matches 52; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1368 tattaattagttattcttctcttgagacagagtcactgtgtgcccaggtc 1427
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3814 TTTTCTTTTCTTTTCTTTTCTTTTGTAGACAGGGCTCTCATTTATGCTAGGCTAGTCT 3755

QY 1428 caaacttgcggtc 1440
||||| ||||| |||
Db 3754 CAAACTTCTGGCC 3742

RESULT 14
US-09-115-954-1/c
; Sequence 1, Application US/09115954B
; Patent No. 6200776
; GENERAL INFORMATION:
; APPLICANT: Boron, Walter F
; APPLICANT: Brill, Antoine M
; APPLICANT: Khandoudi, Nassirah
; APPLICANT: Martin, Xavier
; APPLICANT: Jupe, Steven C
; APPLICANT: Rawlings, Christopher J
; APPLICANT: Doe, Trudy R
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GH-30409
; CURRENT APPLICATION NUMBER: US/09/115,954B
; CURRENT FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: EP97401714.7
; EARLIER FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3912
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-115-954-1

Query Match
Best Local Similarity 1.7%; Score 39.4; DB 4; Length 3912;
Matches 52; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1368 tattaattagttattcttctcttgagacagagtcactgtgtgcccaggtc 1427
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3884 TTTTCTTTTCTTTTCTTTTCTTTTGTAGACAGGGCTCTCATTTATGCTAGGCTAGTCT 3825

QY 1428 caaacttgcggtc 1440
||||| ||||| |||
Db 3824 CAAACTTCTGGCC 3812

RESULT 15
US-08-756-506-1/c
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 13:26:36 ; Search time 4579.66 Seconds
(without alignments)
4749.483 Million cell updates/sec

Title: US-09-445-201-1-copy_8260_10560
Perfect score: 2301
Sequence: 1 tgaataagatgaggtgccc.....aattgctgaagaggcctttt 2301

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1:	gb_est1.*
2:	gb_est2.*
3:	gb_est3.*
4:	gb_est4.*
5:	gb_est5.*
6:	gb_est6.*
7:	gb_est7.*
8:	gb_est8.*
9:	gb_est9.*
10:	gb_est10.*
11:	gb_est11.*
12:	gb_est12.*
13:	gb_est13.*
14:	gb_est14.*
15:	gb_est15.*
16:	gb_est16.*
17:	gb_est17.*
18:	gb_est18.*
19:	gb_est19.*
20:	gb_est20.*
21:	gb_est21.*
22:	gb_est22.*
23:	gb_est23.*
24:	gb_est24.*
25:	gb_est33.*
26:	gb_est34.*
27:	gb_est35.*
28:	gb_est36.*
29:	gb_est37.*
30:	gb_est38.*
31:	gb_est39.*
32:	gb_est40.*
33:	em_estba.*
34:	em_estfun.*
35:	em_esthum1.*
36:	em_esthum2.*
37:	em_esthum3.*
38:	em_esthum4.*
39:	em_esthum5.*
40:	em_esthum6.*
41:	em_esthum7.*
42:	em_esthum8.*
43:	em_esthum9.*

44:	em_esthum10.*
45:	em_esthum11.*
46:	em_esthum12.*
47:	em_esthum13.*
48:	em_esthum14.*
49:	em_esthum15.*
50:	em_esthum16.*
51:	em_esthum17.*
52:	em_esthum18.*
53:	em_esthum19.*
54:	em_esthum20.*
55:	em_esthum21.*
56:	em_esthum22.*
57:	em_esthum23.*
58:	em_esthum24.*
59:	em_esthum25.*
60:	em_esthum26.*
61:	em_esthum27.*
62:	em_esthum28.*
63:	em_estin1.*
64:	em_estin2.*
65:	em_estin3.*
66:	em_estin4.*
67:	em_estin5.*
68:	em_estom1.*
69:	em_estom2.*
70:	em_estov1.*
71:	em_estov2.*
72:	em_estpl1.*
73:	em_estpl2.*
74:	em_estpl3.*
75:	em_estpl4.*
76:	em_estpl5.*
77:	em_estpl6.*
78:	em_estpl7.*
79:	em_estpl8.*
80:	em_estpl9.*
81:	em_estpl10.*
82:	em_estro1.*
83:	em_estro2.*
84:	em_estro3.*
85:	em_estro4.*
86:	em_estro5.*
87:	em_estro6.*
88:	em_estro7.*
89:	em_estro8.*
90:	em_estro9.*
91:	em_estro10.*
92:	em_estro11.*
93:	em_estro12.*
94:	em_estro13.*
95:	em_estro14.*
96:	em_estro15.*
97:	em_estro16.*
98:	em_estro17.*
99:	em_estro18.*
100:	em_estro19.*
101:	em_estro20.*
102:	gb_est25.*
103:	gb_est26.*
104:	gb_est27.*
105:	gb_est28.*
106:	gb_est29.*
107:	gb_est30.*
108:	gb_est31.*
109:	gb_est32.*
110:	gb_est41.*
111:	gb_est42.*
112:	gb_est43.*
113:	gb_est44.*
114:	gb_est45.*
115:	gb_est46.*
116:	gb_est47.*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
1	707	30.7	746	250	AZ815965	AZ815965 2M0084D06
2	560.8	24.4	586	242	AZ364007	AZ364007 1M0110F03
3	79.2	3.4	404	224	AQ126020	AQ126020 HS 3032.B
4	78.6	3.4	464	227	AQ332669	AQ332669 HS 5004.A
5	58.2	2.5	556	245	AZ508851	AZ508851 1M0351G04
6	56.6	2.5	506	248	AZ709930	AZ709930 RPI-24-6
7	55	2.4	256	138	BE651040	BE651040 UI-M-BH3-
8	55	2.4	257	116	AW490147	AW490147 UI-M-BH3-
9	55	2.4	652	117	AW551517	AW551517 1M008A10-
10	54.8	2.4	559	243	AZ423401	AZ423401 1M0202C02
11	53.6	2.3	432	148	BF391061	BF391061 UI-R-CA1-
12	53.4	2.3	320	102	A1838610	A1838610 UI-M-A00-
13	53.4	2.3	429	23	A1661049	A1661049 mz78f09.x
14	53.4	2.3	431	102	A1849973	A1849973 UI-M-BG0-
15	53.4	2.3	470	147	BF320748	BF320748 uz55h11.y
16	53.4	2.3	478	2	AA119334	AA119334 mp80e10.r
17	53.4	2.3	631	121	AW826111	AW826111 us21h04.x
18	53	2.3	569	241	AZ306979	AZ306979 1M0008H17
19	52.4	2.3	311	109	AV102206	AV102206 AV102206
20	52.4	2.3	652	243	AZ403474	AZ403474 1M0171B14
21	52.2	2.3	212	251	AZ913824	AZ913824 RPI-24-1
22	52	2.3	419	247	AZ637662	AZ637662 1M0497J12
23	51.8	2.3	295	109	AV096924	AV096924 AV096924
24	51.6	2.2	458	227	AQ355903	AQ355903 CITBI-E1-
25	51.2	2.2	584	248	AZ750042	AZ750042 RPI-24-8
26	51.2	2.2	621	248	AZ744458	AZ744458 RPI-24-1
27	51.2	2.2	631	243	AZ410081	AZ410081 1M0182G10
28	51.2	2.2	713	173	BG076348	BG076348 H3158F08
29	50.8	2.2	446	17	AI171758	AI171758 EST217740
30	50.6	2.2	609	249	AZ791331	AZ791331 2M0041O10
31	50.4	2.2	311	247	AZ662311	AZ662311 1M0541N17
32	50.2	2.2	367	138	BE655196	BE655196 UI-M-BG1-
33	50.2	2.2	376	241	AZ318225	AZ318225 1M0037H07
34	50.2	2.2	384	4	AA261578	AA261578 mz87a12.r
35	50.2	2.2	505	138	BE689422	BE689422 uw54d12.y
36	50.2	2.2	568	148	BF401678	BF401678 UI-R-CA0-
37	50.2	2.2	578	249	AZ762191	AZ762191 1M0556L23
38	50.2	2.2	632	144	BF121097	BF121097 601757929
39	50.2	2.2	651	244	AZ461771	AZ461771 1M0267H18
40	50	2.2	489	1	AA004129	AA004129 mg82f03.r
41	50	2.2	631	243	AZ431614	AZ431614 1M0216G07
42	49.8	2.2	396	15	A1044824	A1044824 UI-R-CL-k
43	49.6	2.2	575	243	AZ393030	AZ393030 1M0155J19
44	49.6	2.2	624	246	AZ615574	AZ615574 1M0445D12
45	49.4	2.1	264	238	AZ100588	AZ100588 RPI-23-2

ALIGNMENTS

RESULT 1
 AZ815965
 LOCUS AZ815965 746 bp DNA GSS 20-FEB-2001
 DEFINITION 2M0084D06R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC2M0084D06 R, DNA sequence.
 ACCESSION AZ815965
 VERSION
 KEYWORDS
 SOURCE GSS.
 ORGANISM house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 746)
 DUNN,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
 and Wright,D.,Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0084 row: D column: 06
 Seq primer: CACACAGGAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 746.
 FEATURES
 Location/Qualifiers
 1..746
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0084D06"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PWD42 (g114732114g1AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."
 BASE COUNT 162 a 159 c 169 g 256 t
 ORIGIN
 QY 1083 acacatgtacatcttattgaattgggtattatcagggtgagcaggtgagcattgca 1142
 |||||
 Db 1 ACACATGTACATCTTATTAGTGGCTATTTACAGGTGGCATAGCTAGCTATGCA 60
 QY 1143 catctgtgtgtgagccagtgaggaactgctcctgctgctcctcagaaagccacagtg 1202
 |||||
 Db 61 CATCTGTGTAGGTGAGCCAGTGGACAGCTGCCT-CTGGCTGTCTCAGAGGCCACAGTG 119
 QY 1203 tcacggcattggctatttgccttggctctttgctaaactttattgacatggcctcatct 1262
 |||||
 Db 120 TCAGGCGATTGGCTATTTGCTTGGCTCTTTGCTAATACTATTATTCATGCTCATCT 179
 QY 1263 tcgttcacattcacttatttggcccaaacagtcgaatgcagcagctgagcccttagagatcat 1322
 |||||
 Db 180 TCGTTACAGCTTCACTTATTTGGCCCAACAGCTCAATGCCAGCTGAGGCCCTTAGGAGTCAT 239
 QY 1323 ctgtctttagtcagtcgcaattagaaagcctgagtcctgctgctgctatttaattagttatt 1382
 |||||
 Db 240 CTGTTCTTAGTCAGTGTGAATAGAAAGCCTGGATGCTGCTGCTCTAATTAATAGTATT 299
 QY 1383 cttctcttgagacagagatctcaactgtgtgcccaggagtagtctcaaaacttgcggtcca 1442

```

|||||
Db 300 CTTCTCTCTGAGACAGAGTCTCACTGTGTGCGCCAGGCTAGTCTCAAACTTGCCTGCTCA 359
QY 1443 ttgtctcaactcatcagaatgctggctccagggtgtgtgaccacacactagtagctgcgc 1502
Db 360 TTTGCTCTCACTCATCAGAATGCTGGCTTCCAGGTTGTCGACACACACTAGTAGCTGCG 419
QY 1503 gtttaagctaaagctggaagatcctctgatgtcctttaccattggtgggaatgttacagt 1562
Db 420 GTTTTAAGCTAAGAGCTGAAGATCCTGATGCTCTTTTACCATGTTGGGCTGCTTACAGGT 479
QY 1563 tagttgactaaactagttatctgcgtgttaataagctgagtggtgtatgtatctctca 1622
Db 480 TAGTGTACTGAAGAACTAGTTATCTGCTGTGTAAATGACCTGAGTGGTATGTATCTCTCA 539
QY 1623 agatgcttttttgcatttcaatcagttaggttaacaaagt-tcttaagctctcagactgtgta 1681
Db 540 AGATGCTTTTGTGATTTCAATCAGTTAGGTAAACAAGTCTTATGTCTCCAGCTTTGTA 599
QY 1682 ttgcatgagctcagagcttggtatgaatagttggtggacccctagctattgtctcattaga 1741
Db 600 TTGTATGAGCTCAGAGCTTTGATTAATGAGTTGGGACCCCTAGCTATTGCTCAATTAGA 659
QY 1742 cttaacatttttagtttgcctgagtttataaataatgcatgcatgataaacttggg 1801
Db -660 CTTACACTATTTTAGTTTCTCTGAGTTTATGAATATGATGATGATGATGATGATGATG 719
QY 1802 agatatttttctcccaattccctttt 1828
Db 720 AGATATTCTCTCCCAATTCCTTTT 746

RESULT 2
AZ364007 586 bp DNA GSS 02-OCT-2000
LOCUS LM0110F03F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0110F03 F, DNA sequence.
ACCESSION AZ364007
VERSION AZ364007.1 GI:10477707
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
1 (bases 1 to 586)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0110 row: F column: 03
Seq primer: CGTTGTAAACACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 586.
Location/Qualifiers
1. .586
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0110F03"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"

FEATURES
Source

```

```

/lab_host="E: Coli strain XL10-Gold, T1-resistant, F-
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 137 a 125 c 137 g 186 t 1 others
ORIGIN

Query Match 24.4%; Score 560.8; DB 242; Length 586;
Best Local Similarity 98.5%; Pred. No. 2.7e-135;
Matches 576; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 1414 gccaggcctagctcaaaacttgcgtccatttgcctcaactcagatgctgggcttc 1473
Db 2 GGGCCAGCTAGTCTCAAACTTGGCGTCCATTGTCTCACTCATCAGAAATGCTGGGCTTCC 61

QY 1474 aggtgtgtcaccacactaggtagctgcgcttttaagctaagagctgaagatcctgatg 1533
Db 62 AGGTGTGTGCGACACACTAGGTAGCTCGCTTTTAACTAAGAGCTGGAAGATCCTGATG 121

QY 1534 tctttaccatggtgggcatgttacaggttagttgactgaaactagttatctcgtctgtg 1593
Db 122 TCGTTTACCATGTTGGGGCATGTTACAGGTAGTTGACTGAAACTAGTATTCTCGCTGTG 181

QY 1594 taatgacctgagtgatgtatctctcaagatgctttttgcatttcaatcagttaggt 1653
Db 182 TAATGACCTGCGAGTGGTATGTATCTCTCAAGATGCTTTTTCATTTCAATCAGTTAGGT 241

QY 1654 aacaagt-tcttaagctctccagcttggtattggcatgagctcagagcttttgatlaagag 1712
Db 242 AACAAAGTGTCTTATGCTCCAGCTTTGTATTGTTATGAGCTCAGAGCTTTGATTAATGAG 301

QY 1713 ttggacccccctagctattgctcattagactacactatttttagttttgctcgtagttt 1772
Db 302 TTGGGACCCCTAGCTATTGCTCATTAGACTTACACTATTATTAGTTTGTCTGAGTTT 361

QY 1773 atgaatatcatgtatgcatgaacttggagatatttttctcccaattccttttctc 1832
Db 362 ATGAATATCATGTATGCGATGAACCTTGGGAGATATTTCCTCTCCCAATTTCTTTTCTC 421

QY 1833 catttaaatgtctgtcttttagaagccactgctcagcttctcagctcagataccacaaag 1892
Db 422 CATTTAAATGTCTGCTTTTAGAAGCCACTGCTCAGCTTCTGCAGCTCAGATACCAAG 481

QY 1893 gaagtctgtacacagcatgataaaagacaatggcggtgtcacagtggtccctcc 1952
Db 482 GAAGTCTGTGTACACAGCATGATAAAAGACAATGGGACGGNGTACAGTGTGCTCCGTC 541

QY 1953 ttccaggggtatggagacagctgttagagatgtctccagggag 1997
Db 542 TTTCAGGGGTATGGACGAGCTGTAGAGAGATGTCTCCAGGGAG 586

RESULT 3
LOCUS AQ126020/c
DEFINITION HS_3032_B2_C02_MF CIT Approved Human Genomic Sperm Library D Homo

```


sapiens genomic clone Plate=3032 Col=4 Row=F, DNA sequence.

ACCESSION AQ126020
VERSION AQ126020.1 GI:3503186
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 404)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3032 row: F column: 4
Class: BAC ends
High quality sequence stop: 404.

FEATURES
Source Location/Qualifiers
1..404
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_plate=3032 Col=4 Row=F"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
 E-Coli DH10B"

BASE COUNT 160 a 69 c 83 g 92 t
ORIGIN

Query Match 3.4%; Score 79.2; DB 224; Length 404;
Best Local Similarity 56.8%; Pred. No. 8.4e-10;
Matches 135; Conservative 0; Mismatches 53; Indels 14; Gaps 1;

QY 1195 ccacagtgtcacggcattggtctatttgccttggtccttttgcctaacttat----- 1246
Db 316 CCTCAGATTCAATGCCTGACTGGTGTTCGACTGGGTATTTACTTATCTCTCTATTAT 257

QY 1247 -----tgacatggccatctctcggtcacgttcaacttatgtgccacaacgtcaatgc 1300
Db 256 TCGGTCTCATACTTCGGTCCTATTGTTTCATACTCTCTATTTCGCCAAGGTCMAATGC 197

QY 1301 cagctgaggccttaggagtcatctgttcttagtcagtcgaattagaacctggatgcc 1360
Db 196 CAGTTTAGGATGATGGGAGTCATTTTTTCTTAGTGATGACTAGAAAGCTTTGGGTGCC 137

QY 1361 tgctgctattaattagttatt 1382
Db 136 TGCCCAACATCAATTACTTTTT 115

RESULT 4
AQ332669/c

LOCUS AQ332669 464 bp DNA GSS 06-MAR-1999
DEFINITION HS_5004_A1_A09.SP8E.SPC111 Human Male BAC Library Homo sapiens
genomic clone Plate=580 Col=17 Row=A, DNA sequence.

ACCESSION AQ332669
VERSION AQ332669
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 173 a 162 c 89 g 135 t
ORIGIN

Query Match 2.4%; Score 54.8; DB 243; Length 559;
Best Local Similarity 59.7%; Pred. No. 0.0021;
Matches 92; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1330 tctcagtcgcaattagaaagcctgagcctgcctcctctattatgattctctct 1389
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 558 TAGTGTGTCAGACAGATGATGTTGGTGGTGGTGGTGGTGGTGGTGGTGGT 499
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1390 tctgagacagagctcactgtgtgcccaggtgagtcctcaaaactgggtccattgtct 1449
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 498 TTTGAGGACAGGTTTTCGTGTGTAGCCAGCTGGCTGTAACTTGCAGATCCGCTCT 439
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1450 catcatcagaatcgtgggcttcacaggtgtgtgc 1483
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 438 CAGATTTCATGAGTGGTGGGAATACAGATGTGTAC 405
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
BF391061 432 bp mRNA EST 27-NOV-2000
LOCUS
DEFINITION UI-R-CAL-bcf-f-12-0-UI.s1 UI-R-CAL Rattus norvegicus cDNA clone
ACCESSION BF391061
VERSION BF391061.1 GI:11375908
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 432)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized medulla library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research

Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 16-133, >PB1D10\$INE/Alu
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
Source

1. .432
Location/Qualifiers
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CAL-bcf-f-12-0-UI"
/clone_lib="UI-R-CAL"

/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; The UI-R-CAL library is a subtracted library derived from the following tissues: thalamus, cerebellum, hypothalamus, medulla, pons, midbrain, cerebral cortex, corpus striatum, testis, and hippocampus. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB="UI-R-CAL"
TAG_TISSUE="medulla"
TAG_SEQ="GAACCG"

BASE COUNT 115 a 85 c 79 g 153 t
ORIGIN

Query Match 2.3%; Score 53.6; DB 148; Length 432;
Best Local Similarity 66.4%; Pred. No. 0.004;
Matches 77; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 1368 tattaattagttattctctctcagacagagtcactgtgtgcccaggtgtgtct 1427
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 5 TTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 64
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1428 caaacttcggtccattgtctcactcatcagaatgctgggctccaggtgtgtgc 1483
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 65 GGAACCTGCTACTCTACTGTTCCAACTCCAGAGCTGGGACTGAAGGTGTTTC 120
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
AI838610
LOCUS
DEFINITION UI-M-A00-abz-f-12-0-UI.s1 NIH BMAP.MPG Mus musculus cDNA clone
ACCESSION AI838610
VERSION AI838610.1 GI:5472823
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 320)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: m5t@mail.nih.gov

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized pineal glands library cDNA Library Preparation: M.B.

Soares Lab Clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined. The following repetitive elements were found in this cDNA sequence: 1-44, >AT-rich#Low_complexity 103-204, >PBID10#SINE/Alu

Seq primer: M13 Forward
POLYA=Yes.

FEATURES

Source

Location/Qualifiers

1. 320

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UI-M-A00-abz-f-12-0-UI"

/clone_lib="NIH_BMAP_MPG"

/dev_stage="27-32 days"

/lab_host="DH10B (Life Technologies)"

/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_MPG library is a non-normalized library constructed from mouse pineal gland. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Ronaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories.

TAG_LIB=NIH_BMAP_MPG

TAG_TISSUE=pineal-glands

TAG_SEQ=CGAC"

BASE COUNT 69 a 81 c 63 g 107 t

ORIGIN

Query Match 2.3%; Score 53.4; DB 102; Length 320;

Best Local Similarity 69.9%; Pred. No. 0.0042;

Matches 72; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 1385 tctcttgagacagagctcactgtgtgcccaggttagctcaaaacttgcggtccatt 1444

||||| ||||||| || ||||| ||||||| || ||||||| ||||| |||||

Db 105 TCCTCTTTGAGACAGGAACTTGTCTATAGCCAGGCTGGCTTTGAACCTGGGATCCCTCT 164

||||| ||||| || ||||| ||||| ||||| ||||| ||||| |||||

QY 1445 tgtctcactcatcagaatgctgggtccaggtgtgtgcacca 1487

||||| ||||| || ||||| ||||| ||||| ||||| ||||| |||||

Db 165 TGGCTCAGTTTCCAACTTCCAGGGTTCAGGTGTGCCCCCA 207

||||| ||||| || ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13

AI661049

LOCUS

DEFINITION

mz78f09.x1 Soares mouse NML Mus musculus cDNA clone IMAGE:719561 3' similar to SW:RFXA_HUMAN 000287 REGULATORY FACTOR X-ASSOCIATED

PROTEIN ; mRNA sequence.

ACCESSION

AI661049

VERSION

AI661049.1 GI:4764632

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 429)

Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,

Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person

,B., Waller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter

,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,

Waterston,R. and Wilson,R.

The WashU-NCI Mouse EST Project 1999

Unpublished (1999)

Other ESTs: mz78f09.y1

Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

FEATURES

Source

Location/Qualifiers

1. 429

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:719561"

/clone_lib="Soares mouse NML"

/tissue_type="liver"

/lab_host="DH10B"

/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTACCATCTGAAGTGGAGCGCGGAATCTTTTTCCTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. Library constructed and normalized by Bento Soares and M.Patima Bonaldo."

BASE COUNT 98 a 105 c 89 g 136 t 1 others

ORIGIN

Query Match 2.3%; Score 53.4; DB 23; Length 429;

Best Local Similarity 69.9%; Pred. No. 0.0045;

Matches 72; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 1385 tctcttgagacagagctcactgtgtgcccaggttagctcaaaacttgcggtccatt 1444

||||| ||||||| || ||||| ||||||| || ||||||| ||||| |||||

Db 89 TCCTCTTTGAGACAGGAACTTGTCTATAGCCAGGCTGGCTTTGAACCTGGGATCCCTCT 148

||||| ||||| || ||||| ||||| ||||| ||||| ||||| |||||

QY 1445 tgtctcactcatcagaatgctgggtccaggtgtgtgcacca 1487

||||| ||||| || ||||| ||||| ||||| ||||| ||||| |||||

Db 149 TGGCTCAGTTTCCAACTTCCAGGGTTCAGGTGTGCCCCCA 191

||||| ||||| || ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14

AI849973

LOCUS

DEFINITION

UI-M-BG0-aib-b-06-0-UI-s1 NIH_BMAP MSC Mus musculus cDNA clone

UI-M-BG0-aib-b-06-0-UI 3', mRNA sequence.

ACCESSION

AI849973

VERSION

AI849973.1 GI:5493879

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 431)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

9704477

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: mEST@mail.nih.gov

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dT track served to verify it as a clone from the

non-normalized spinal cord library cDNA Library Preparation: M.B.

Soares Lab Clone distribution: NIH BMAP cDNA clones will be made

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
Possible reversed clone: similarity on wrong strand
High quality sequence stop: 321.

available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined. The following repetitive elements were found in this cDNA sequence: 1-41. >AT_richLow complexity 100-201, >PB1D10#SINE/Alu
Seq primer: M13 Forward
POLYA=yes.

FEATURES

source Location/Qualifiers

1. .431
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BG0-aib-b-06-0-UI"
/clone_lib="NIH_BMAP_MSC"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_MSC library is a non-normalized library constructed from mouse spinal cord. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories.
TAG_LIB=NIH_BMAP_MSC
TAG_TISSUE=spinal-cord
TAG_SEQ=TCGAA"

BASE COUNT 97 a 105 c 84 g 145 t

ORIGIN

Query Match 2.3%; Score 53.4; DB 102; Length 431;
Best Local Similarity 69.9%; Pred. No. 0.0045;
Matches 72; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
Qy 1385 tctcttgagacagagtcctcactgtgtgcccaggtagtctcaaaacttgcggtccatt 1444
Db 102 TCTCTTTTGAGACAGGAACCTTGCTATGAGCCAGGCTGGCTTTGAACTTGGGATCCTCT 161
Qy 1445 tgtctcactcagaatgctgggtccagggtgtgcacca 1487
Db 162 TGGCTCAGTTTCCAAAGTTCAGGGGTTCCAGGTGTGCCCCCAA 204

RESULT 15

BF320748/c

LOCUS

BF320748 470 bp mRNA EST 29-DEC-2000
uz55h11.y1 NCI-CGAP_Mam6 Mus musculus cDNA clone IMAGE:3673029 5'
similar to SW:RFXA_HUMAN 000287 REGULATORY FACTOR X-ASSOCIATED
PROTEIN ; mRNA sequence.

ACCESSION

BF320748

VERSION

BF320748.1

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 470)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-r@mail.nih.gov

Tissue Procurement: Jeffrey Green M.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml

MGI:1433797

High quality sequence stop: 453.

FEATURES

source

1. .470
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3673029"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 139 a 91 c 110 g 130 t

ORIGIN

Query Match 2.3%; Score 53.4; DB 147; Length 470;
Best Local Similarity 69.9%; Pred. No. 0.0046;
Matches 72; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
Qy 1385 tctcttgagacagagtcctcactgtgtgcccaggtagtctcaaaacttgcggtccatt 1444
Db 304 TCTCTTTTGAGACAGGAACCTTGCTATGAGCCAGGCTGGCTTTGAACTTGGGATCCTCT 245
Qy 1445 tgtctcactcagaatgctgggtccagggtgtgcacca 1487
Db 244 TGGCTCAGTTTCCAAAGTTCAGGGGTTCCAGGTGTGCCCCCAA 202

Search completed: June 21, 2001, 15:33:53

Job time: 7637 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 18:05:44 ; Search time 150.46 Seconds
(without alignments)
2798.729 Million cell updates/sec

Title: US-09-445-201-1_COPY_8336_10608

Perfect score: 2273

Sequence: 1 tctcgagacagacgcgtgt.....agggttcctattgatcc 2273

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*

2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*

3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*

4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*

5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47.6	2.1	7218	1	US-08-232-463-14
2	45.6	2.0	51259	3	US-08-781-891-209
3	43.2	1.9	152331	4	US-09-128-155-16
4	42.2	1.9	14507	3	US-08-785-150-1
5	40.8	1.8	152331	4	US-09-128-155-16
6	40.4	1.8	176373	4	US-09-128-155-17
7	40.2	1.8	3860	1	US-08-126-587C-5
8	40	1.8	1920	1	US-08-087-772A-1
9	39.6	1.7	246240	2	US-08-724-394A-20
10	39.6	1.7	246240	2	US-08-724-394A-21
11	39.6	1.7	246240	2	US-08-724-394A-22
12	39.4	1.7	2880	4	US-09-115-954-3
13	39.4	1.7	3842	4	US-09-115-954-7
14	39.4	1.7	3912	4	US-09-115-954-1
15	39.2	1.7	11725	2	US-08-756-506-1
16	39.2	1.7	17041	1	US-08-076-011-1
17	39	1.7	35060	3	US-08-814-095-7
18	38.6	1.7	2908	3	US-08-487-799-1
19	38.6	1.7	6769	1	US-08-480-784-20
20	38.6	1.7	6769	1	US-08-483-553-20
21	38.6	1.7	6769	1	US-08-487-002-20
22	38.6	1.7	6769	1	US-08-483-554B-20
23	38.6	1.7	6769	4	US-08-488-011B-20
24	38.6	1.7	6769	4	US-08-850-727-20
25	38.6	1.7	6769	5	PCT-US95-10202-20
26	38.6	1.7	6769	5	PCT-US95-10203-20
27	38.6	1.7	6769	5	PCT-US95-10220-20

c	28	38.4	1.7	501	3	US-08-699-628-1	Sequence 1, Appl
	29	38.4	1.7	2520	2	US-08-454-557C-50	Sequence 50, Appl
	30	38.4	1.7	2520	2	US-08-340-426D-50	Sequence 50, Appl
	31	38.4	1.7	2520	2	US-08-450-673C-50	Sequence 50, Appl
	32	38.4	1.7	2520	5	PCT-US95-17111A-50	Sequence 50, Appl
c	33	38.2	1.7	5095	1	US-08-092-817-3	Sequence 3, Appl
	34	38	1.7	31571	1	US-08-323-443B-1	Sequence 1, Appl
c	35	38	1.7	40352	3	US-08-846-111D-15	Sequence 15, Appl
	36	38	1.7	53526	3	US-08-658-136-2	Sequence 2, Appl
	37	38	1.7	53577	3	US-08-658-136-1	Sequence 1, Appl
	38	37.6	1.7	615	4	US-08-998-416-186	Sequence 186, App
	39	37.4	1.6	7720	4	US-09-318-448-5	Sequence 5, Appl
	40	37	1.6	4084	3	US-08-866-340-1	Sequence 1, Appl
	41	37	1.6	4460	4	US-09-103-875-4	Sequence 4, Appl
	42	36.8	1.6	4922	2	US-08-330-272-5	Sequence 5, Appl
	43	36.8	1.6	4922	5	PCT-US95-13663-5	Sequence 5, Appl
c	44	36.6	1.6	5375	4	US-08-757-223-7	Sequence 7, Appl
c	45	36.6	1.6	10825	3	US-08-652-265-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

```

; INFORMATION FOR SEQ ID NO: 209:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51259 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-781-891-209

Query Match          2.08;   Score 45.6; DB 3;   Length 51259;
Best Local Similarity 62.1%;   Pred. No. 0.0095;
Matches 72; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 1292 tattaatagttattctctcttgagacagagtcctcaactgtgtgcccaggctagtct 1351
      ||||| ||| ||| ||||| ||||| ||||| ||||| ||| ||| |||
Db 23743 TTTTAGTTTAGTTTACTTTTTTTAGACAGGGTCTCAGCTGTAGCTGGGACAGCT 23802
      ||||| ||| ||| ||||| ||||| ||||| ||||| ||| ||| |||

QY 1352 caaactgcggtccattgtctcaactcatcagaatgtcgggcttccagggtgtgac 1407
      ||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 23803 CCACCCCTGTCCCTTTTCTCTCACCCCTCCTGAGTGTCTGGGATCACAGGCGTGTGC 23858
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
US-09-128-155-16
US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

```

```

; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match      1.9%; Score 43.2; DB 4; Length 152331;
Best Local Similarity 63.5%; Pred. No. 0.095;
Matches 66; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 1261 tgcgaattagaagcgtggatgcctgcctctatttaattgtattctctctctgaga 1320
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 142531 tgcattgcacaggcgtggttttttttttttttttttttttttttttttttttcgtagaga 142590

QY 1321 cagatctcaactgtgtggccaggctagctcaaacctgcggtc 1364
      | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 142591 ctgggtctctctgtattggccaggctagctcgaacctctgggc 142634

RESULT 4
US-08-785-150-1
; Sequence 1, Application US/08785150
; Patent No. 6027915
; GENERAL INFORMATION:
; APPLICANT: Morris, Arvia E.
; APPLICANT: Lee, Chi-Chang
; APPLICANT: Thomas, James N.
; TITLE OF INVENTION: Expression Augmenting Sequence Elements
; Patent No. 6027915

```


GENERAL INFORMATION:
APPLICANT: Hayden, Michael
APPLICANT: Goldberg, Paul
APPLICANT: Andrew, Susan
APPLICANT: Rommens, Johanna M.
APPLICANT: Lin, Biaoyang
TITLE OF INVENTION: Process for Isolating Genes and the Gene
TITLE OF INVENTION: Causative of Huntington's Disease and Differential 3'
TITLE OF INVENTION: Polyadenylation in the Gene
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell Seltzer Park & Gibson
STREET: 1211 E. Morehead Street
CITY: Charlotte
STATE: NO. 5534438th Carolina
COUNTRY: USA
ZIP: 28234
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/126,587C
FILING DATE: 24-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Layton Jr., Samuel G.
REGISTRATION NUMBER: 22,807
REFERENCE/DOCKET NUMBER: 3477-84
TELECOMMUNICATION INFORMATION:
TELEPHONE: 704-377-1561
TELEFAX: 704-334-2014
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 380 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 4p16.3
UNITS: bp
US-08-126-587C-5

Query Match 1.8%; Score 40.2; DB 1; Length 380;
Best Local Similarity 60.8%; Pred. No. 0.02;
Matches 66; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
QY 1292 tattaatagttattctctcttgagacagagtgctcactgtgtgcccaggtactgtct 1351
Db 339 TTTTITTTTTTTTTTTTTTTTGAGACGGAGCTGTCTCTGCGCCAGCGGACT 280
QY 1352 caaacttgcggtccattgtctcactcatcagaatgctgggctccagg 1400
Db 279 GCGGACTGCAGTGGCGCAATCTCGGCTCACTGAAAGCTCCGCTCCCGG 231

RESULT 8
US-08-087-772A-1/c
Sequence 1, Application US/0808772A
Patent No. 5691155
GENERAL INFORMATION:
APPLICANT: Namias, Clara
APPLICANT: Emorine, Jean L.
APPLICANT: Strosberg, Donny A.
TITLE OF INVENTION: Nucleotide Sequences Encoding the Murine
TITLE OF INVENTION: Beta3-Adrenergic Receptor and Their Applications
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:

ADDRESSEE: Bell, Seltzer, Park & Gibson
STREET: Post Office Drawer 34009
CITY: Charlotte
STATE: No. 5691155th Carolina
COUNTRY: USA
ZIP: 28234
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/087,772A
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Linker, Raymond O.
REGISTRATION NUMBER: 26,419
REFERENCE/DOCKET NUMBER: 3339-195
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-881-3140
TELEFAX: 919-881-3175
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1920 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 568..1731
US-08-087-772A-1

Query Match 1.8%; Score 40; DB 1; Length 1920;
Best Local Similarity 63.5%; Pred. No. 0.061;
Matches 61; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
QY 1306 tcttctcttgagacagagtgctcactgtgtgcccaggtactgtctcaaaacttgcggtcc 1365
Db 97 TCTTATTTGTTGAGATAGTCTTCTATATGTTGCCAGATTTGCTGGACCTCCAGATCC 38
QY 1366 attgtctcactcatcagaatgctgggctccagg 1401
Db 37 TTCTGCTCAACCTCCCGAGTGTGGATTACAGAT 2

RESULT 9
US-08-724-394A-20
Sequence 20, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
TITLE OF INVENTION: Sequences and Antibodies Thereof
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs

```

: APPLICANT: Jupe, Steven C
: APPLICANT: Rawlings, Christopher J
: APPLICANT: Doe, Trudy R
: TITLE OF INVENTION: NOVEL COMPOUNDS
: FILE REFERENCE: GH-30409
: CURRENT APPLICATION NUMBER: US/09/115,954B
: CURRENT FILING DATE: 1998-07-15
: EARLIER APPLICATION NUMBER: EP97401714.7
: EARLIER FILING DATE: 1997-07-16
:

```

```

; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7
; LENGTH: 3842
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-115-954-7

Query Match      1.7%; Score 39.4; DB 4; Length 3842;
Best Local Similarity 71.2%; Pred. NO. 0.14;
Matches 52; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```

```

Db      3814  TTTTTTTTTTTTTTTTTTTTTTTCTAGACAGGGTCTCATTATATTGCCTAGGCTAGTCT 3755
QY      1352  caaaccttcgggc 1364
Db      3754  CAAACTTCGGCC 3742

RESULT 14
US-09-115-954-1/c
; Sequence 1, Application US/09115954B
; Patent No. 6200776
; GENERAL INFORMATION:
; APPLICANT: Boron, Walter F
; APPLICANT: Brill, Antoine M
; APPLICANT: Khandoudi, Nassirah
; APPLICANT: Martin, Xavier
; APPLICANT: Jupe, Steven C
; APPLICANT: Rawlings, Christopher J
; APPLICANT: Doe, Trudy R
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GH-30409
; CURRENT APPLICATION NUMBER: US/09/115,954B
; CURRENT FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: EP97401714.7
; EARLIER FILING DATE: 1997-07-15

```

```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3912
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-115-954-1

Query Match      1.7%; Score 39.4; DB 4; Length 3912;
Best Local Similarity 71.2%; Pred. No. 0.14;
Matches 52; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```

Db	3884	TTTTTTTTTTTTTTTTTTTGTAGACAGGGTCTCATTATATTCCTAGGCTAGTCT	3825
Qy	1352	caaaacttcggtc	1364
Db	3824	CAAACTCTCGCC	3812

RESULT 15
S-08-756-506-1/c

Search completed: June 21, 2001, 18:20:53
Job time: 17446 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 15:33:53 ; Search time 4579.66 Seconds
(without alignments)
4691.688 Million cell updates/sec

Title: US-09-445-201-1-COPY_8336_10608
Perfect score: 2273
Sequence: 1 tctcgagacagacgcgtgt.....aggcttctctattgatcc 2273

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1:	gb_estl1:
2:	gb_est2:
3:	gb_est3:
4:	gb_est4:
5:	gb_est5:
6:	gb_est6:
7:	gb_est7:
8:	gb_est8:
9:	gb_est9:
10:	gb_est10:
11:	gb_est11:
12:	gb_est12:
13:	gb_est13:
14:	gb_est14:
15:	gb_est15:
16:	gb_est16:
17:	gb_est17:
18:	gb_est18:
19:	gb_est19:
20:	gb_est20:
21:	gb_est21:
22:	gb_est22:
23:	gb_est23:
24:	gb_est24:
25:	gb_est33:
26:	gb_est34:
27:	gb_est35:
28:	gb_est36:
29:	gb_est37:
30:	gb_est38:
31:	gb_est39:
32:	gb_est40:
33:	em_estba:
34:	em_estfun:
35:	em_esthum1:
36:	em_esthum2:
37:	em_esthum3:
38:	em_esthum4:
39:	em_esthum5:
40:	em_esthum6:
41:	em_esthum7:
42:	em_esthum8:
43:	em_esthum9:

44:	em_esthum10:
45:	em_esthum11:
46:	em_esthum12:
47:	em_esthum13:
48:	em_esthum14:
49:	em_esthum15:
50:	em_esthum16:
51:	em_esthum17:
52:	em_esthum18:
53:	em_esthum19:
54:	em_esthum20:
55:	em_esthum21:
56:	em_esthum22:
57:	em_esthum23:
58:	em_esthum24:
59:	em_esthum25:
60:	em_esthum26:
61:	em_esthum27:
62:	em_esthum28:
63:	em_estin1:
64:	em_estin2:
65:	em_estin3:
66:	em_estin4:
67:	em_estin5:
68:	em_estom1:
69:	em_estom2:
70:	em_estov1:
71:	em_estov2:
72:	em_estpl1:
73:	em_estpl2:
74:	em_estpl3:
75:	em_estpl4:
76:	em_estpl5:
77:	em_estpl6:
78:	em_estpl7:
79:	em_estpl8:
80:	em_estpl9:
81:	em_estpl10:
82:	em_estro1:
83:	em_estro2:
84:	em_estro3:
85:	em_estro4:
86:	em_estro5:
87:	em_estro6:
88:	em_estro7:
89:	em_estro8:
90:	em_estro9:
91:	em_estro10:
92:	em_estro11:
93:	em_estro12:
94:	em_estro13:
95:	em_estro14:
96:	em_estro15:
97:	em_estro16:
98:	em_estro17:
99:	em_estro18:
100:	em_estro19:
101:	em_estro20:
102:	gb_est25:
103:	gb_est26:
104:	gb_est27:
105:	gb_est28:
106:	gb_est29:
107:	gb_est30:
108:	gb_est31:
109:	gb_est32:
110:	gb_est41:
111:	gb_est42:
112:	gb_est43:
113:	gb_est44:
114:	gb_est45:
115:	gb_est46:
116:	gb_est47:

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

```

```

190: gb_est110:*
191: gb_est111:*
192: gb_est112:*
193: gb_est113:*
194: gb_est114:*
195: gb_est115:*
196: gb_est116:*
197: gb_est117:*
198: gb_est118:*
199: gb_est119:*
200: gb_est120:*
201: gb_est121:*
202: gb_est122:*
203: gb_est123:*
204: gb_est124:*
205: gb_est125:*
206: gb_est126:*
207: gb_est127:*
208: gb_est128:*
209: gb_est129:*
210: gb_est130:*
211: gb_est131:*
212: gb_est132:*
213: gb_est133:*
214: gb_est134:*
215: gb_est135:*
216: gb_est136:*
217: gb_est137:*
218: gb_est138:*
219: gb_est139:*
220: gb_est140:*
221: gb_est141:*
222: gb_est142:*
223: gb_est143:*
224: gb_est144:*
225: gb_est145:*
226: gb_est146:*
227: gb_est147:*
228: gb_est148:*
229: gb_est149:*
230: gb_est150:*
231: gb_est151:*
232: gb_est152:*
233: gb_est153:*
234: gb_est154:*
235: gb_est155:*
236: gb_est156:*
237: gb_est157:*
238: gb_est158:*
239: gb_est159:*
240: gb_est160:*
241: gb_est161:*
242: gb_est162:*
243: gb_est163:*
244: gb_est164:*
245: gb_est165:*
246: gb_est166:*
247: gb_est167:*
248: gb_est168:*
249: gb_est169:*
250: gb_est170:*
251: gb_est171:*
252: gb_est172:*
253: gb_est173:*
254: gb_est174:*
255: gb_est175:*
256: gb_est176:*
257: gb_est177:*
258: gb_est178:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.


```

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes=Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (gi|4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      137 a  125 c  137 g  186 t      1 others
ORIGIN

Query Match      24.7%; Score 560.8; DB 242; Length 586;
Best Local Similarity 98.5%; Pred. No. 9.7e-135;
Matches 576; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 1338 gccacggctagtcctcaaacctgcggctccatttgtctcactcaccagaatgcctggcttcc 1397
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2 GCCCCAGCTAGTCTCAAACTTCGGGTCCTATTTCTCTACTCATCAGAAATGCTGGGCTTCC 61

QY 1398 aggtgtgtgaccacactaggtagctcgctgttaagctaaagctagaagatcctgatg 1457
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 AGGTGTGTGCACCACACTAGGTAGCTCGCGGTTTTAGCTAAGAGCTGGAAGATCCTGATG 121

QY 1458 tccttaaccatggtggcattgattacaggttagttgactgaaacctagttatctcgctgtg 1517
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 122 TCCCTTTACATGTGGGATGTTACAGGTTAGTTGACTGAAACTAGTTATCTCGCGTGTG 181

QY 1518 taatgacctgcagtgtatgtatctctcaagatgcttttgcattccaatcagttaggt 1577
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 182 TAATGACCTGCAGTGGTATGTATCTCTCAAGATGCTTTTGTGATTTTCAATCAGTTAGGT 241

QY 1578 acaagt-tcttaagctccagcttggtattggcatgagctcagagctttgattaaatgag 1636
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 242 AACAAAGTCTTATGTCTCCAGCTTGTATTGGTATGAGCTCAGAGCTTTGATTAAATGAG 301

QY 1637 ttggagcccccctagctattgtcctattagacttacactatttttagttttgctcgtgatt 1696
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 302 TTGGGACCCCTAGCTATTGTCTATTAGACTTACACTATTTTAGTTTTCCTCTGAGTTT 361

QY 1697 atgaatatgcatgatgaacttggagagatattttcttcccacattccttttctc 1756
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 362 ATGAATATGCAATGATGCATGAACCTGGGAGATATTCTCTCCCAATTCCTTTTCCCTC 421

QY 1757 catttaaatgtgctgtctttagaagccactgcctcagctcttcgcagctcagataccacaa 1816
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 422 CATTTAAATGTGCTCTTTTAGAGCCACTGCCTCAGCTTCGACGCTCAGATACCAAAAG 481

QY 1817 gaagctgtgtacagacatgataaagacaatgggacggggtcacagtggtgcctccccc 1876
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 482 GAAAGTCTGGTACACAGCATGATAAAGACAATGGGACGGNGTCACAGTGGCTCCCGTCCC 541

QY 1877 ttccaggggtatgagacgaactgtagagatgtctccagggag 1921
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 542 TTTCAGGGGTATGAGACGAGCTGTAGAGAGATGTCTCCAGGGAG 586

RESULT 3
AQ126020/c
LOCUS
DEFINITION HS_3032_B2_C02_Mf Clt Approved Human Genomic Sperm Library D Homo
23-SEP-1998
GSS
DNA
404 bp

```

sapiens genomic clone Plate=3032 Col=4 Row=F, DNA sequence.
 AQ126020
 VERSION AQ126020.1 GI:3503186
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 404)
 AUTHORS Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T.,
 Kellar A., Shaker R., Furlong J., Young J., Zhao S., Adams M.D. and
 Hood L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 3032 row: F column: 4
 Class: BAC ends
 High quality sequence stop: 404.
 FEATURES
 source
 1..404
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=3032 Col=4 Row=F"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelBAC11; BAC Clones in
 E-Coli DH10B"
 BASE COUNT 160 a 69 c 83 g 92 t
 ORIGIN
 Query Match 3.5%; Score 79.2; DB 224; Length 404;
 Best Local Similarity 66.8%; Pred. No. 9.7e-10;
 Matches 135; Conservative 0; Mismatches 53; Indels 14; Gaps 1;
 QY 1119 ccacagtgtcacgcatgtgctatttgccttgctcttgccttaactttat-----1170
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 316 CCTCAGATTCAATGCACCTGGATTTTGCACCTGGATTTTACTATATCTCTCTATTAT 257
 QY 1171 -----tgacatggcctcactcttcgttcacgttcacttattgcccacacacgtcaatgc 1224
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 256 TCCGTCATACCTCGCTCTATTGTTTCATCTCTCTATTGTTCCACGCAAGTCAATGC 197
 QY 1225 cagctgagccttagagtcactctgttcgttcagtcgaattagaaagcctgagtc 1284
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 196 CAGTTAGGCAATAGGAGGTCATTTTCTTAGTTGATGATGACTTAGAAGCTGGGTGCC 137
 QY 1285 tgctgtgctatttaattagttatt 1306
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 136 TGCCCAACATCAATTACTTTT 115
 RESULT 4
 AQ332669/c
 LOCUS HS_5004.A1.A09.SP6E.RPC111 Human Male BAC Library Homo sapiens
 DEFINITION genomic clone Plate=580 Col=17 Row=A, DNA sequence.
 ACCESSION AQ332669
 VERSION AQ332669.2 GI:4341028
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 464)
 AUTHORS Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T.,
 Kellar A., Shaker R., Furlong J., Young J., Zhao S., Adams M.D. and
 Hood L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT On Feb 19, 1999 this sequence version replaced gi:4130296.
 Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 580 row: A column: 17
 Seq primer: SP6
 Class: BAC ends
 High quality sequence stop: 464.
 FEATURES
 source
 1..464
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=580 Col=17 Row=A"
 /clone_lib="RPC111 Human Male BAC Library"
 /sex="male"
 /cell_type="Lymphocytes"
 /note="Vector: pBAC3.6; RPC111 Human Male BAC Library"
 BASE COUNT 178 a 75 c 93 g 113 t
 ORIGIN
 Query Match 3.5%; Score 78.6; DB 227; Length 464;
 Best Local Similarity 66.5%; Pred. No. 1.4e-09;
 Matches 125; Conservative 0; Mismatches 62; Indels 1; Gaps 1;
 QY 1120 cacagtgtcacgcatgtgctatttgccttgctcttgccttaactttat-tgacatgg 1178
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 298 CACTGATGTGTCACCTGGTATTACTTACTTACTCTCTCTATTTCCTCTCATCTT 239
 QY 1179 cctcattctcgttcacgttcacttattgcccacacacgtcaatgcacgtgagcctta 1238
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 238 CGTCCTATTGTTTCATCTCTCTATTGTCACGCAAGTCAATGCCAGTTTAGGCCTAG 179
 QY 1239 gagtcactctgttccttcagtcagtcgaattagaaagcctgagtcgctgctgtattat 1298
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 178 GGAGTCATTTTCTTAGTTGATGATGACTTAGAAGCTTGGTGCNTGCCCCCATCATCAT 119
 QY 1299 tagttatt 1306
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 118 TACTTTT 111
 RESULT 5
 AZ508851/c
 LOCUS AZ508851 556 bp DNA GSS 05-OCT-2000
 DEFINITION 1M0351G04R Mouse 10kb plasmid UUGCLM library Mus musculus genomic
 clone UUGCLM0351G04 R, DNA sequence.
 ACCESSION AZ508851
 VERSION AZ508851.1 GI:10690167
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 556)
 AUTHORS Dunn D., Aoyagi A., Barber M., Beacorn T., Duval B., Hamill C.,
 Islam H., Longacre S., Mahmoud M., Meenen E., Pedersen T., Reilly
 M., Rose M., Rose R., Stokes R., Tingey A., von Niederhausern A.
 and Wright D., Weiss R.

Contact: Shay
Department of
The Institute
9712 Medical
Tel: 301 838
Fax: 301 838

8
Tel: 301 838
Fax: 301 838
Email: szhao@
Clones are d
library avail
(pdejong@mai
Resources (ht
page: <http://>
Plate: 67 ro
Seq primer: 1
Class: HAC en

Query Match	Best Local Similarity	Matches 109; Conservation
1258	cagtcgaattagaa	
324	CAGTGAATCCAGCAAA	
1318	agacagagtcctact	
384	AAAGAAGATCTTACT	
1378	catcagaatctggg	
444	CTCTGAGTCTGGG	

DEFINITION	UI-M-BH3-asy	UI-M-BH3-asy-
ACCESSION	BE651040	
VERSION	BE651040.1	G
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
	Eukaryota; Me	
	Mammalia; Eut	

Eukaryota; Me
Mammalia; Eut

//RAD_HOSC- E. COLI STRAIN XL10-GOLD, TI-RESISTANT, F-"

//RAD_HOSC- E. COLI STRAIN XL10-GOLD, TI-RESISTANT, F-"

available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined. The following repetitive elements were found in this cDNA sequence: 1-41, >AT-rich/Low-complexity 100-201, >PB1D10#SINE/Alu
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

Location/Qualifiers
1..431
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BG0-aib-b-06-0-UI"
/clone_lib="NIH_BMAP_MSC"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3P-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_MSC library is a non-normalized library constructed from mouse spinal cord. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories.
TAG_LIB="NIH_BMAP_MSC"
TAG_TISSUE="spinal-cord"
TAG_SEQ="TCGAA"

BASE COUNT 97 a 105 c 84 g 145 t

ORIGIN

Query Match 2.3%; Score 53.4; DB 102; Length 431;
Best Local Similarity 69.9%; Pred. No. 0.0049;
Matches 72; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 1309 tctctctgagacagagtctcactgtgtgcccaggctagctcaaaccttgcgtccatt 1368
Db 102 TCTCTTTTGAGACAGGAAGTTGCTATGTAGCCAGGCTGGCTTTGAACCTGGGATCCTCT 161
QY 1369 tatctcactcatcaagaatcgtggctccagggtgtgtgcacca 1411
Db 162 TGGCTCAGTTTTCAGAGTTCCAGGGGTTCACAGGTGTGCCCCCAA 204

RESULT 15

BF320748/c
LOCUS BF320748 470 bp mRNA EST 29-DEC-2000
DEFINITION uz55h11.y1 NCI-CGAP_Mam6 Mus musculus cDNA clone IMAGE:3673029 5' similar to SW:RFXA_HUMAN O00287 REGULATORY FACTOR X-ASSOCIATED PROTEIN ; mRNA sequence.
ACCESSION BF320748
VERSION BF320748.1 GI:11269760
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 470)
REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
TITLE Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml

MGI:1433797

High quality sequence stop: 453.

FEATURES

Location/Qualifiers
1..470
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3673029"
/clone_lib="NCI-CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 139 a 91 c 110 g 130 t

ORIGIN

Query Match 2.3%; Score 53.4; DB 147; Length 470;
Best Local Similarity 69.9%; Pred. No. 0.005;
Matches 72; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 1309 tctctctgagacagagtctcactgtgtgcccaggctagctcaaaccttgcgtccatt 1368
Db 304 TCTCTTTTGAGACAGGAAGTTGCTATGTAGCCAGGCTGGCTTTGAACCTGGGATCCTCT 245
QY 1369 tgtctcactcatcaagaatcgtggctccagggtgtgtgcacca 1411
Db 244 TGGCTCAGTTTTCAGAGTTCCAGGGGTTCACAGGTGTGCCCCCAA 202

Search completed: June 21, 2001, 15:33:59
Job time: 7643 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 18:32:40 ; Search time 8015.43 Seconds
(without alignments)
993.820 Million cell updates/sec

Title: US-09-445-201-1_COPY_10094_10608
Perfect score: 515
Sequence: 1 tttaaatgtgtctgtcttag.....agggtcttctattggtatcc 515

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues 2688314
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_ba3.*
4: gb_in1.*
5: gb_in2.*
6: gb_in3.*
7: gb_om.*
8: gb_ov.*
9: gb_pat1.*
10: gb_pat2.*
11: gb_ph.*
12: gb_pl1.*
13: gb_pl2.*
14: gb_pl3.*
15: gb_pl4.*
16: em_ba1.*
17: em_ba2.*
18: em_fun.*
19: em_hugo_hum.*
20: em_htgo_inv.*
21: em_htgo_rod.*
22: em_htg_hum1.*
23: em_htg_hum2.*
24: em_htg_hum3.*
25: em_htg_hum4.*
26: em_htg_hum5.*
27: em_htg_hum6.*
28: em_htg_hum7.*
29: em_htg_hum8.*
30: em_htg_inv1.*
31: em_htg_inv2.*
32: em_htg_inv3.*
33: em_htg_rod.*
34: em_hum1.*
35: em_hum2.*
36: em_hum3.*
37: em_hum4.*
38: em_hum5.*
39: em_hum6.*
40: em_hum7.*
41: em_in.*
42: em_om.*
43: em_or.*

44: em_ov.*
45: em_pat.*
46: em_ph.*
47: em_pl.*
48: em_ro.*
49: em_sts.*
50: em_sy.*
51: em_un.*
52: em_vi.*
53: gb_sts1.*
54: gb_sts2.*
55: gb_sts3.*
56: gb_sy.*
57: gb_un.*
58: gb_vil.*
59: gb_vil2.*
60: gb_hgl.*
61: gb_hgl2.*
62: gb_hgl3.*
63: gb_hgl4.*
64: gb_hgl5.*
65: gb_hgl6.*
66: gb_hgl7.*
67: gb_hgl8.*
68: gb_hgl9.*
69: gb_hgl10.*
70: gb_hgl11.*
71: gb_hgl12.*
72: gb_hgl13.*
73: gb_hgl14.*
74: gb_hgl15.*
75: gb_hgl16.*
76: gb_hgl17.*
77: gb_hgl18.*
78: gb_hgl19.*
79: gb_hgl20.*
80: gb_hgl21.*
81: gb_hgl22.*
82: gb_hgl23.*
83: gb_hgl24.*
84: gb_hgl25.*
85: gb_pr1.*
86: gb_pr2.*
87: gb_pr3.*
88: gb_pr4.*
89: gb_pr5.*
90: gb_pr6.*
91: gb_pr7.*
92: gb_pr8.*
93: gb_pr9.*
94: gb_rol.*
95: gb_rol2.*
96: gb_in4.*
97: gb_pr10.*
98: em_ba3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	515	100.0	12845	9	AX002124 Sequence
2	510	99.0	510	94	AF061804 Mus muscu
3	430	83.5	430	94	AF153058 Mus muscu
C 4	304.6	59.1	135552	65	AC019251 Homo sapi
C 5	304.6	59.1	214348	66	AC021220 Homo sapi
C 6	254.4	49.4	126330	63	AC013745 Homo sapi
C 7	39.4	7.7	847	13	AF327424 Arabidops
C 8	39.4	7.7	8785	12	AC007659 Arabidops

9 39.4 7.7 196646 80 AL365329 Mus muscu
 10 39.4 7.7 259487 80 AL365322 Mus muscu
 11 38.8 7.5 162299 75 AC079086 Homo sapi
 12 38.8 7.5 194038 87 AC010103 Homo sapi
 13 37 7.2 157721 83 AP003265 Oryza sat
 14 37 7.2 182615 78 AF217246 Homo sapi
 15 36.8 7.1 148267 81 AL390758 Homo sapi
 16 36.4 7.1 99858 93 HSJ1059A9 Human DNA
 17 36.4 7.1 187584 79 AL162731 Homo sapi
 18 36.4 7.1 189352 62 AC011967 Homo sapi
 19 36.4 7.1 202081 90 AL391987 Human DNA
 20 36.4 7.1 215283 62 AC011966 Homo sapi
 21 35.8 7.0 144075 63 AC013763 Homo sapi
 22 35.6 6.9 176961 63 AC013356 Homo sapi
 23 34.6 6.8 914 53 CNS01FJT Anophelis
 24 34.8 6.8 141970 89 AL138761 Human DNA
 25 34.8 6.8 149340 66 AC021514 Homo sapi
 26 34.8 6.8 170909 80 AL360170 Homo sapi
 27 34.6 6.7 32501 93 HSAP2B Homo sapi
 28 34.6 6.7 133965 88 AF064861 Homo sapi
 29 34.6 6.7 171648 77 AC087103 Papio ham
 30 34.6 6.7 190349 87 AC009475 Homo sapi
 31 34.6 6.7 191904 61 AC009956 Homo sapi
 32 34.6 6.7 340000 92 HS21C079 Homo sapi
 33 34.4 6.7 183575 82 AL590127 Homo sapi
 34 34.4 6.7 184084 70 AC026659 Homo sapi
 35 34.4 6.7 220242 64 AC016759 Homo sapi
 36 34.2 6.6 4552 5 AF064258 Strongylo
 37 34.2 6.6 125487 93 HSDJ685D6 Homo sapi
 38 34.2 6.6 168074 75 AC079069 Homo sapi
 39 34.2 6.6 185393 86 AC006552 Homo sapi
 40 34 6.6 169845 65 AC013054 Homo sapi
 41 34 6.6 175380 78 AC091045 Homo sapi
 42 34 6.6 191466 73 AC068355 Homo sapi
 43 33.8 6.6 140691 89 AF159056 Homo sapi
 44 33.8 6.6 179137 82 AF000853 Homo sapi
 45 33.8 6.6 217218 61 AC009132 Homo sapi

ALIGNMENTS

RESULT 1
 LOCUS AX002124 12845 bp DNA PAT 10-MAR-2000
 DEFINITION Sequence 1 from Patent WO9855638.
 ACCESSION AX002124
 VERSION AX002124.1 GI:7241839
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unclassified.
 REFERENCE 1 (bases 1 to 12845)
 AUTHORS Breier,G. and Risau,W.
 TITLE REGULATORY SEQUENCES CAPABLE OF CONFERRING EXPRESSION OF A HETEROLOGOUS DNA SEQUENCE IN ENDOTHELIAL CELLS IN VIVO AND USES THEREOF
 JOURNAL Patent: WO 9855638-A 1 10-DEC-1998;
 FEATURES
 source Location/Qualifiers
 1..12845 /organism="unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 3321 a 2893 c 2863 g 3723 t 45 others
 ORIGIN
 Query Match 100.0%; Score 515; DB 9; Length 12845;
 Best Local Similarity 100.0%; Pred. No. 3.3e-154;
 Matches 515; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 tttaaattgctgtcttttagaagcaactcctcagcttctgcagctcagatccaagga 60
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 10094 TTAAATGCTGCTGCTTTAGAAAGCCACTGCCTCAGCTTCTGCAGCTCAGATACCAAGGA 10153
 Oy 61 agctgggtacacagcatgataaaacacaaatggaggggtcacagtggtcccgctccctt 120
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 10154 AGTGTGTACACAGCATGATAAAACACAAATGGACGGGTACAGTGGCTCCGTCCTT 10213
 Oy 121 tcaggggtatggagacagcgtgtagagagatgtctccagggaggttttcattaatcagcaa 180
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 10214 TCAGGGGTATGGACGACGAGCTGTAGAGAGATGCTCCAGGGAGTTCATTATACGCAA 10273
 Oy 181 tttaagtcagatctgtgcactctatcttacaagaatgtcagtggtggcctgagatcatca 240
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 10274 TTTAGTCAGATCTGTGCATCTCTATCTTTTACAAGAAATGTCAGTGGGCTGAGATCATCA 10333
 Oy 241 gatggaggttcattcggttttcaaatgtccgtatctcttttgaagacctgaagttggcaa 300
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 10334 GATGAGGTTTCATCGGTTTCAATGTCCTCCGTATCTTTTGAAGACCTTGAAGTTGGCAA 10393
 Oy 301 cgcgggaaaaacagaaactccacctcgtggtcgtgtaattgcagagctgtgtgtgtgttg 360
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 10394 CGCAGGAAAACAGGAACCTCCACCTGGTCCCTGTAATTCGACAGAGCTGTGTGTGTGTG 10453
 Oy 361 tgaccatctgccattctctctgttatcacagagctgtgaaactttaactggactgggg 420
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 10454 TGACCATCTGCCATCTCTCTGTATGACAGAGCTTGTAACCTGGACTGGGG 10513
 Oy 421 caaagtcaatccacacctttatataaatgaattgctgaagagggccttttaaaacttgagtg 480
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 10514 CAAAGTCAATCCACCTTTATACAATGAATGCTGAAGAGGCTTTTAAACTTGGAGTG 10573
 Oy 481 tgcattgttatgaagggcttctctatttgatgc 515
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 10574 TGCATTGTTTATGAAGGCGCTTCTCTATTGGATCC 10608
 RESULT 2
 AF061804 510 bp DNA ROD 23-AUG-1999
 LOCUS Mus musculus fetal liver kinase 1 (Flkl) gene, intronic enhancer.
 DEFINITION AF061804
 ACCESSION AF061804
 VERSION AF061804.1 GI:5757676
 KEYWORDS
 SOURCE house mouse.
 ORGANISM Mus musculus
 REFERENCE 1 (bases 1 to 510)
 AUTHORS Kappel,A., Ronicke,V., Damert,A., Flamme,I., Risau,W. and Breier,G.
 TITLE Identification of vascular endothelial growth factor (VEGF) receptor-2 (Flk-1) promoter/enhancer sequences sufficient for angioblast and endothelial cell-specific transcription in transgenic mice
 JOURNAL Blood 93 (12), 4284-4292 (1999)
 MEDLINE 95290783
 PUBMED 10361126
 REFERENCE 2 (bases 1 to 510)
 AUTHORS Kappel,A., Ronicke,V., Damert,A., Flamme,I., Risau,W. and Breier,G.
 TITLE Direct Submission
 JOURNAL Submitted (28-APR-1998) Molekulare Zellbiologie, Max Planck Institut fuer physiologische und klinische Forschung, Parkstrasse 1, Bad Nauheim 61231, Germany
 FEATURES
 source Location/Qualifiers
 1..510 /organism="Mus musculus"
 /strain="C57BL"
 /db_xref="taxon:10090"
 1..510 /gene="Flkl"
 /note="fetal liver kinase 1"
 1..510 /gene="Flkl"
 /note="intronic"
 enhancer

BASE COUNT 129 a 105 c 131 g 145 t
ORIGIN

Query Match 99.0%; Score 510; DB 94; Length 510;
Best Local Similarity 100.0%; Pred. No. 8.7e-153;
Matches 510; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 taaatgtctgtcttttagaagccactgcctcagctctctcagctcagatcaccacaaaggaag 62
|||||
Db 1 TAAATGTCTGTCTTTAGAGCCACTGCCCTCAGCTTCTGCAGCTCAGATACCAAGGAAG 60
|||||
QY 63 tctgttacacagcatgataaaagacaaatgggacggggtcacagtggtccctcttc 122
|||||
Db 61 TCTGGTACACAGCATGATAAAGACAAATGGACGGGTACACAGTGGCTCCCGTCCCTTTC 120
|||||
QY 123 aggggtatggagacgagcttaagagatgtctccagggagttttcattaatcagcaatt 182
|||||
Db 121 AGGGGTATGGAGACGAGCTGTAGAGAGATGCTCCAGGGAGTTTTTCATTAAATCAGCAATT 180
|||||
QY 183 tagtcagatctgtgcctcctctatctttacaagaaatgtcagtgggcctgagatcaccaga 242
|||||
Db 181 TAGTCAGATCTGTGCATCTATCTTTACAAGAAATGTCACTGGGCTGAGATCATCAGA 240
|||||
QY 243 tggagttcatcggtttcaatgtcccgatccctgtatccttttgaagaccttgagttggcaacg 302
|||||
Db 241 TGGAGGTTTCATCGGGTTTCAATGTCCCGTATCCTTTTGAAGACCTTGAAGTTGGCAACG 300
|||||
QY 303 caggaaacaggaactccaccctgggtgcggtgaattgcagagctgtgtgtttgttg 362
|||||
Db 301 CAGGAACAGGAACATCCACCCCTGGTGGCGGTGAATTCGACAGCTGTGTGTGGTTGTG 360
|||||
QY 363 accatctgcccattcttcctgttatgacagagcttgtgaactttaactggagctgggca 422
|||||
Db 361 ACCATCTGCCATCTTCTCTGTATGACAGAGCTTGTGAACCTTAACTGGGACTGGGCA 420
|||||
QY 423 aagtcacccacactttatatacaatgactgctgaagagccttttaaaacttgagtg 482
|||||
Db 421 AAGTCAATCCACCTTTATACAATGAATTTGCTGAAGAGGCGCTTTTAAACTTGGAGTGTG 480
|||||
QY 483 cattgttatggagggcttctcattgga 512
|||||
Db 481 CATGTGTTATGGAGGGCTTTCCTATTGGA 510
|||||

RESULT 3

AF153058 430 bp DNA ROD 22-MAY-2000
LOCUS Mus musculus tyrosine kinase FLK-1 minimal enhancer sequence.
DEFINITION AF153058
ACCESSION AF153058.1 GI:7963625
VERSION
KEYWORDS house mouse.
SOURCE
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 430)
Kappel,A., Risau,W. and Breier,G.
prerequisite role of SCL/tal-1, GATA and Ets transcription factor
binding sites for the in vivo function of Flk-1 gene regulatory
elements
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 430)
AUTHORS Kappel,A., Risau,W. and Breier,G.
TITLE Direct Submission
JOURNAL Submitted (21-MAY-1999) Molecular Cell Biology, Max-Planck,
Institut fuer Physiologische und Klinische Forschung, Parkstrasse
1, Bad Nauheim 61231, Germany
FEATURES
Location/Qualifiers
1..430
/organism="Mus musculus"
/strain="129Sv"
/db_xref="taxon:10090"

enhancer
BASE COUNT 108 a 94 c 112 g 116 t
ORIGIN

/lissue_type="liver"
/dev_stage="fetus"
1..430
/note="tyrosine kinase FLK-1 minimal enhancer"

Query Match 83.5%; Score 430; DB 94; Length 430;
Best Local Similarity 100.0%; Pred. No. 4.5e-127;
Matches 430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 aatgtgtctgtcttttagaagccactgcctcagctctctcagctcagatcaccacaaaggaag 63
|||||
Db 1 AAATGTCTGTCTTTTACAAGCCACTGCTCAGCTTCTGCAGCTCAGATACCAAGGAAGT 60
|||||
QY 64 ctggtcacacagcatgataaaagacaatgggacggggtcacagtggtccctcttca 123
|||||
Db 61 CTGGTACACAGCATGATAAAGACAAATGGACGGGTACACAGTGGCTCCCGTCCCTTTCA 120
|||||
QY 124 ggggtatggagacgagctgtagagagatgtctccagggagtttttcattatcagaattt 183
|||||
Db 121 GGGGTATGGAGACGAGCTGTAGAGAGATGCTCCAGGGAGTTTTTCATTAATCAGCAATT 180
|||||
QY 184 agtcagatctgtgcctcctatgtctttacaagaaatgtcagtgggcctgagatcatcagat 243
|||||
Db 181 AGTCAGATCTGTGCATCTATGCTTTTACAAGAAATGTCACTGGGCTGAGATCATCAGAT 240
|||||
QY 244 ggggttcatcggttttcaatgtccctgtatccttttgaagaccttgagttggcaacgc 303
|||||
Db 241 GGAGGTTCAATCGGGTTTCAATGTCCCGTATCCTTTTGAAGACCTTGAAGTTGGCAACG 300
|||||
QY 304 aggaacacaggaactccaccctgggtgcggtgaattgcagagctgtgtgtgtttgtga 363
|||||
Db 301 AGAAAACAGGAACCTCCACCCTGGTGGCGGTGAATTCGAGAGCTGTGTGTGGTTGTGA 360
|||||
QY 364 ccattgtcccatcttctcctgttatgacagagcttgaactttaactgggactggggcaa 423
|||||
Db 361 CCATCTGCCATCTTCTCTGTATGACAGAGCTTGTGAACCTTTAACTGGGACTGGGCAA 420
|||||
QY 424 agtcaatccc 433
|||||
Db 421 AGTCAATCCC 430
|||||

RESULT 4

AC019251/c 135552 bp DNA HTG 13-JUL-2000
LOCUS Homo sapiens chromosome 4 clone RP11-5E20 map 4, LOW-PASS SEQUENCE
DEFINITION AC019251
ACCESSION AC019251
VERSION AC019251.2 GI:7158903
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 135552)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 4, clone RP11-5E20
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 135552)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G., Castle,A.,
Choepe,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferrelira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,J., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marguis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,

Norman, C.H., O'Connor, T., O'Donnell, P., Olivari, T.M., Peterson, K.,
 Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
 Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
 Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

Direct Submission
 Submitted (31-DEC-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 5, 2000 this sequence version replaced gl:6649482.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project information

Center project name: L2005

Center clone name: 5_E_20

* NOTE: This record contains 153 individual

* sequencing reads that have not been assembled into

* contigs. Runs of N are used to separate the reads

* and the order in which they appear is completely

* arbitrary. Low-pass sequence sampling is useful for

* identifying clones that may be gene-rich and allows

* overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone

* will be sequenced to completion. In the event that

* the record is updated, the accession number will

* be preserved.

1 752: contig of 752 bp in length
 753 852: gap of 100 bp
 853 1637: contig of 785 bp in length
 1638 1737: gap of 100 bp
 1738 2506: contig of 769 bp in length
 2507 2606: gap of 100 bp
 2607 3392: contig of 786 bp in length
 3393 3492: gap of 100 bp
 3493 4261: contig of 769 bp in length
 4262 4361: gap of 100 bp
 4362 5128: contig of 767 bp in length
 5129 5228: gap of 100 bp
 5229 6000: contig of 772 bp in length
 6001 6100: gap of 100 bp
 6101 6856: contig of 756 bp in length
 6857 6956: gap of 100 bp
 6957 7734: contig of 778 bp in length
 7735 7834: gap of 100 bp
 7835 8636: contig of 802 bp in length
 8637 8736: gap of 100 bp
 8737 9526: contig of 790 bp in length
 9527 9626: gap of 100 bp
 9627 10374: contig of 748 bp in length
 10375 10474: gap of 100 bp
 10475 11244: contig of 770 bp in length
 11245 11344: gap of 100 bp
 11345 12128: contig of 784 bp in length
 12129 12228: gap of 100 bp
 12229 13009: contig of 781 bp in length
 13010 13109: gap of 100 bp
 13110 13926: contig of 817 bp in length
 13927 14026: gap of 100 bp
 14027 14758: contig of 732 bp in length
 14759 14858: gap of 100 bp
 14859 15655: contig of 797 bp in length
 15656 15755: gap of 100 bp
 15756 16543: contig of 788 bp in length
 16544 16643: gap of 100 bp
 16644 17436: contig of 793 bp in length
 17437 17536: gap of 100 bp

17537 18309: contig of 773 bp in length
 18310 18409: gap of 100 bp
 18410 19183: contig of 774 bp in length
 19184 19283: gap of 100 bp
 19284 20038: contig of 755 bp in length
 20039 20138: gap of 100 bp
 20139 20878: contig of 740 bp in length
 20879 20978: gap of 100 bp
 20979 21780: contig of 802 bp in length
 21781 21880: gap of 100 bp
 21881 22642: contig of 762 bp in length
 22643 22742: gap of 100 bp
 22743 23549: contig of 807 bp in length
 23550 23649: gap of 100 bp
 23650 24442: contig of 793 bp in length
 24443 24542: gap of 100 bp
 24543 25313: contig of 771 bp in length
 25314 25413: gap of 100 bp
 25414 26164: contig of 751 bp in length
 26165 26264: gap of 100 bp
 26265 27045: contig of 781 bp in length
 27046 27145: gap of 100 bp
 27146 27921: contig of 776 bp in length
 27922 28021: gap of 100 bp
 28022 28812: contig of 791 bp in length
 28813 28912: gap of 100 bp
 28913 29663: contig of 751 bp in length
 29664 29763: gap of 100 bp
 29764 30555: contig of 792 bp in length
 30556 30655: gap of 100 bp
 30656 31427: contig of 772 bp in length
 31428 31527: gap of 100 bp
 31528 32317: contig of 790 bp in length
 32318 32417: gap of 100 bp
 32419 33198: contig of 781 bp in length
 33199 33298: gap of 100 bp
 33299 34081: contig of 783 bp in length
 34082 34181: gap of 100 bp
 34182 34938: contig of 757 bp in length
 34939 35038: gap of 100 bp
 35039 35795: contig of 757 bp in length
 35796 35895: gap of 100 bp
 35896 36624: contig of 729 bp in length
 36625 36724: gap of 100 bp
 36725 37501: contig of 777 bp in length
 37502 37601: gap of 100 bp
 37602 38375: contig of 774 bp in length
 38376 38475: gap of 100 bp
 38476 39264: contig of 789 bp in length
 39265 39364: gap of 100 bp
 39365 40148: contig of 784 bp in length
 40149 40248: gap of 100 bp
 40249 41001: contig of 753 bp in length
 41002 41101: gap of 100 bp
 41102 41871: contig of 770 bp in length
 41872 41971: gap of 100 bp
 41972 42757: contig of 786 bp in length
 42758 42857: gap of 100 bp
 42858 43646: contig of 789 bp in length
 43647 43746: gap of 100 bp
 43747 44507: contig of 761 bp in length
 44508 44607: gap of 100 bp
 44608 45362: contig of 755 bp in length
 45363 45462: gap of 100 bp
 45463 46193: contig of 731 bp in length
 46194 46293: gap of 100 bp
 46294 47090: contig of 797 bp in length
 47091 47190: gap of 100 bp
 47191 47981: contig of 791 bp in length
 47982 48081: gap of 100 bp
 48082 48867: contig of 786 bp in length
 48868 48967: gap of 100 bp
 48968 49724: contig of 757 bp in length


```
* 49725 49824: gap of 100 bp
* 49825 50586: contig of 762 bp in length
* 50587 50686: gap of 100 bp
* 50687 51456: contig of 770 bp in length
* 51457 51556: gap of 100 bp
* 51557 52334: contig of 778 bp in length
* 52335 52434: gap of 100 bp
* 52435 53204: contig of 770 bp in length
* 53205 53304: gap of 100 bp
* 53305 54077: contig of 773 bp in length
* 54078 54177: gap of 100 bp
* 54178 54958: contig of 781 bp in length
* 54959 55058: gap of 100 bp
* 55059 55833: contig of 775 bp in length
* 55834 55933: gap of 100 bp
* 55934 56716: contig of 783 bp in length
* 56717 56816: gap of 100 bp
* 56817 57612: contig of 796 bp in length
* 57613 57712: gap of 100 bp
* 57713 58422: contig of 710 bp in length
* 58423 58522: gap of 100 bp
* 58523 59310: contig of 788 bp in length
* 59311 59410: gap of 100 bp
* 59411 60186: contig of 776 bp in length
* 60187 60286: gap of 100 bp
* 60287 61061: contig of 775 bp in length
* 61062 61161: gap of 100 bp
* 61162 61891: contig of 730 bp in length

Query Match      59.1%; Score 304.6; DB 65; Length 135552;
Best Local Similarity 79.0%; Pred. No. 2e-86;
Matches 418; Conservative 0; Mismatches 94; Indels 17; Gaps 4;

QY 1 tttaaatgtgctgtctttagaagccactgcctcagcttctgcagctcagatcacatacacaagga 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63576 TTTTATGTACACTTTTATGACCCATTTATTCAGCTGTGGCAGTTTGTGTTACACAGGGA 63517

QY 61 agtcgtgtacacagcatgataaaagaacaatgggacggggtcacagtggctcccgctccctt 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63516 AGCGCACTAGAAAA-TTGATAAAGGAAAAATGACAAAGGTTCATAGATTCTCTCACTCCCT 63458

QY 121 tcagggtatggagacagctgtagagatgtctccaggagtttctaatcaatcaagcaa 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63457 TCA-GGGTAGCTAGATGAACATATATAAAATCCGCTTAAGTGGGATTCGTTAATCAGCAA 63399

QY 181 tttagtcagatctgtgcactctatgtcttacaagaaatgcagtgggcc----- 229
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63398 TTTAGTCAATGTGTACATCCTATGTTCTATAGAATGTCAGTGGGTCTTTTCCCAAGG 63339

QY 230 ---tgagatcatcagatggaggttcacatcggtttcaatgtcccgatcctctttgtgaagac 286
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63338 GACTGAGATCATCAGATGAAGGTTTCATTTGGTTTCAATGTCCTGATCCCTTTTGTGAAGAC 63279

QY 287 cttagaagtgtggaacacaggaagaaacaggaactccacccctggtgcggtgaattcagagct 346
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63278 CTTTGAAGTTGGCAATCAGGAAACAGGAACCTCCACCCCTAGTCATGAAATTCAGAACT 63219

QY 347 gttgtgtgtgtgtgacacatcgccacttcttcctgtgtatgacagagctgtggaacttt 406
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63218 GTTGTGTTGTTTATGACCATCTGCCCATCTCTCTGTTTATGACACAGCTTGTGAACCTT 63159

QY 407 aactgggaactgggggcaaaagtcaa-tcccaccttttataaatgaattgctgaagagccctt 465
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63158 TACTGAGAATGGTGAAGAAAGTAAATTTCCACGTTTTTATATAAATGAATTCCTGAAGAGCCTT 63099

QY 466 ttaaaacttggaagtgcattgtttatggaagggcttctcatttgatc 514
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63098 TTAAGATATAGATATGCATTTGTTTATGGAAGGTGTTTCTTATTAGGTC 63050

RESULT 5
AC021220/c LOCUS
```

DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
COMMENT

Homo sapiens chromosome 4 clone RP11-530I17, WORKING DRAFT
SEQUENCE, 38 unordered pieces.
AC021220
AC021220.4 GI:11178158
HTG; HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
Waterston,R.H.
Direct Submission
2 (bases 1 to 214348)
Submitted (15-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Nov 15, 2000 this sequence version replaced gi:9309528.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0530117
----- Summary Statistics -----
Sequencing vector: M13; 87%
Sequencing vector: plasmid; 13%
Chemistry: Dye-primer ET; 87% of reads
Chemistry: Dye-terminator Big Dye; 13% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 195065 bases at least Q40
Consensus quality: 206600 bases at least Q30
Consensus quality: 203756 bases at least Q20
Insert size: 161000; agarose-fp
Insert size: 210648; sum-of-contigs
Quality coverage: 4.96 in Q20 bases; agarose-fp
Quality coverage: 4.62 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 38 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1584: contig of 1584 bp in length
* 1585 1684: gap of unknown length
* 1685 4194: contig of 2510 bp in length
* 4195 4294: gap of unknown length
* 4295 6424: contig of 2130 bp in length
* 6425 9844: gap of unknown length
* 9845 9944: gap of unknown length
* 9945 14101: contig of 4157 bp in length
* 14102 14201: gap of unknown length
* 14202 21279: contig of 7078 bp in length
* 21280 21379: gap of unknown length
* 21380 31346: contig of 9967 bp in length
* 31347 31447: gap of unknown length
* 31447 40526: contig of 9080 bp in length
* 40527 40626: gap of unknown length
* 40627 52862: contig of 12236 bp in length
* 52863 66090: contig of 13128 bp in length
* 66091 66190: gap of unknown length
* 66191 82659: contig of 16459 bp in length
* 82660 82759: gap of unknown length
* 82760 104797: contig of 22038 bp in length
* 104798 104897: gap of unknown length

```

* 104898 129319: contig of 24422 bp in length
* 129320 129419: gap of unknown length
* 176511 176511: contig of 47092 bp in length
* 176612 176611: gap of unknown length
* 177662 177666: contig of 1055 bp in length
* 177667 177667: gap of unknown length
* 177767 178997: contig of 1231 bp in length
* 178998 179097: gap of unknown length
* 179098 180242: contig of 1145 bp in length
* 180243 180342: gap of unknown length
* 180343 181442: contig of 1100 bp in length
* 181443 181542: gap of unknown length
* 181543 182642: contig of 1100 bp in length
* 182643 182742: gap of unknown length
* 182743 184165: contig of 1423 bp in length
* 184166 184265: gap of unknown length
* 184266 185612: contig of 1247 bp in length
* 185613 185612: gap of unknown length
* 185613 187121: contig of 1509 bp in length
* 187122 187221: gap of unknown length
* 187222 188524: contig of 1303 bp in length
* 188525 188624: gap of unknown length
* 188625 189782: contig of 1158 bp in length
* 189783 189882: gap of unknown length
* 189883 191312: contig of 1430 bp in length
* 191313 191412: gap of unknown length
* 191413 192837: contig of 1425 bp in length
* 192838 192937: gap of unknown length
* 192938 194117: contig of 1180 bp in length
* 194118 194217: gap of unknown length
* 194218 195948: contig of 1731 bp in length
* 195949 196048: gap of unknown length
* 196049 197744: contig of 1696 bp in length
* 197745 197844: gap of unknown length
* 197845 199589: contig of 1645 bp in length
* 199590 201304: contig of 1715 bp in length
* 201305 201404: gap of unknown length
* 201405 203316: contig of 1912 bp in length
* 203317 203416: gap of unknown length
* 203417 204490: contig of 1074 bp in length
* 204491 204590: gap of unknown length
* 204591 206676: contig of 2086 bp in length
* 206677 206776: gap of unknown length
* 206777 208603: contig of 2027 bp in length
* 208604 208903: gap of unknown length
* 208904 210597: contig of 1694 bp in length
* 210598 210697: gap of unknown length
* 210698 212097: contig of 1400 bp in length
* 212098 212197: gap of unknown length
* 212198 214348: contig of 2151 bp in length.
FEATURES
    Location/Qualifiers
        1..214348
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="4"
            /clone="RP11-530I17"
        1..1584
            /note="assembly_name:Contig100"
            1685..4194
                /note="assembly_name:Contig101"
            4295..5424
                /note="assembly_name:Contig102"
            5525..9844
                /note="assembly_name:Contig103"
            9945..14101
                /note="assembly_name:Contig104"
            14202..21279
                /note="assembly_name:Contig105"
            21380..31346
                /note="assembly_name:Contig106"
            31447..40526
                /note="assembly_name:Contig107

```

```

clone_end:T7
vector_side:left
40627..52862
/note="assembly_name:Contig108"
52963..66090
/note="assembly_name:Contig109"
66191..82659
/note="assembly_name:Contig110"
82760..104797
/note="assembly_name:Contig111"
104898..129319
/note="assembly_name:Contig112
clone_end:SP6
vector_side:right
129420..176511
/note="assembly_name:Contig113"
176612..177666
/note="assembly_name:Contig133"
177767..178997
/note="assembly_name:Contig137"
179098..180242
/note="assembly_name:Contig156"
180343..181442
/note="assembly_name:Contig159"
181543..182642
/note="assembly_name:Contig166"
182743..184165
/note="assembly_name:Contig169"
184266..185512
/note="assembly_name:Contig175"
185613..187121
/note="assembly_name:Contig177"
187222..188524
/note="assembly_name:Contig178"
188625..189782
/note="assembly_name:Contig180"
189883..191312
/note="assembly_name:Contig185"
191413..192837
/note="assembly_name:Contig186"
192938..194117
/note="assembly_name:Contig187"
194218..195948
/note="assembly_name:Contig188"
196049..197744
/note="assembly_name:Contig189"
197845..199489
/note="assembly_name:Contig190"
199590..201304
/note="assembly_name:Contig191"
201405..203316
/note="assembly_name:Contig193"
203417..204490

```

Query Match 59.1%; Score 304.6; DB 66; Length 214348;
 Best Local Similarity 79.0%; Pred. No. 2.2e-86;
 Matches 418; Conservative 0; Mismatches 94; Indels 17; Gaps 4;

```

Qy 1 ttttaagtgtctttagaagccactgcctcagcttctcagctcagatcagatacacaagga 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 100740 TTTTAAATGTCAGTCTTTATGAGCCATATTTTACGCTGGCAGTTGGTTACCAAGGGA 100681

Qy 61 agtctggtacacagcatgataaagacaaatggagcagggggtcagctggtccgcgtccctt 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 100680 AGCGCACTAGAAAA-TTGATAAGGAAATGAGACAAGGTATAGATTCTCTCACTCCCT 100622

Qy 121 tcaggggtatggacagcagctgtagagagatgtctccaggggagtttcttaatacagcaa 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 100621 TCA-GGGTACGTAGATGAACATATAAAAAATCCGCTAAGTGGGATTCGTTTATACAGCA 100563

Qy 181 tttagtcagatctgtgcctcctctcttaccagaagaatgtcagtgaggcc----- 229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 100562 TTTAGTCAAAATGTGTACATCCTCTATGTTCTTATAAGAAATGTCAGTGGGTCTCTTTCCCAAGG 100503

```

```

QY 230 ---tgagatcatcagatgaggttcacgtggtttcaatgtcccgatccttttgaagc 286
|||||
Db 100502 GAGTGAGATCATCAGATGAAGGTTTCATTTGGTTTCAATGTCCTGATCCTTTGTAAGAC 100443
|||||
QY 287 cttgaagtggcaacacagagaaacaggaaactccaccctgggtccgtgaattgcagagct 346
|||||
Db 100442 CTTGAAGTTGGCAATCAGAGAAACAGAACTCCACCCCTAGCTGCTGATGAATTCAGAACT 100383
|||||
QY 347 gttgttggtttgacacatcgtcccatcttccctgttatgacagagcttggaacttt 406
|||||
Db 100382 GTTGTGTTGGTTTATGACCATGTCGCCCATCTCTCTGTTATGACACAGCTTGTAAGACTTT 100323
|||||
QY 407 aactggactgggggcaaaacaa-tcccacctttatcacatgaattgctgaagagcctt 465
|||||
Db 100322 TACTGAGATGTTGGAAGTAAATTCCTAGTTTATACATGAATTCCTGTAAGAGGCTT 100263
|||||
QY 466 ttaaaacttggaagtgcattgtttatggaagggcttctcattgcatc 514
|||||
Db 100262 TTAAGATATAGATGATGTCATTTGTTTATGGAAGGTGTTCTCTATTAGGTC 100214
|||||

RESULT 6
AC013745/c
LOCUS AC013745
DEFINITION Homo sapiens chromosome 4 clone RP11-92P18 map 4, LOW-PASS SEQUENCE
SAMPLING.
AC013745
VERSION AC013745.3 GI:9114527
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 126330)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 4, clone RP11-92P18
2 (bases 1 to 126330)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Balwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Collangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wymann,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6910837.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2025
Center clone name: 92_P_18
-----
* NOTE: This record contains 147 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely

```

```

* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 722: contig of 722 bp in length
* 823 822: gap of 100 bp
* 823 1542: contig of 720 bp in length
* 1543 1642: gap of 100 bp
* 1643 2371: contig of 729 bp in length
* 2372 2471: gap of 100 bp
* 2472 3212: contig of 741 bp in length
* 3213 3312: gap of 100 bp
* 3313 4049: contig of 737 bp in length
* 4050 4149: gap of 100 bp
* 4150 4840: contig of 691 bp in length
* 4841 4940: gap of 100 bp
* 4941 5676: contig of 736 bp in length
* 5677 5776: gap of 100 bp
* 5777 6521: contig of 745 bp in length
* 6522 6621: gap of 100 bp
* 6622 7344: contig of 723 bp in length
* 7345 7444: gap of 100 bp
* 7445 8161: contig of 717 bp in length
* 8162 8261: gap of 100 bp
* 8262 9014: contig of 753 bp in length
* 9015 9114: gap of 100 bp
* 9115 9827: contig of 713 bp in length
* 9828 9927: gap of 100 bp
* 9928 10650: contig of 723 bp in length
* 10651 10750: gap of 100 bp
* 10751 11479: contig of 729 bp in length
* 11480 11579: gap of 100 bp
* 11580 12318: contig of 739 bp in length
* 12319 12418: gap of 100 bp
* 12419 13137: contig of 719 bp in length
* 13138 13237: gap of 100 bp
* 13238 13959: contig of 722 bp in length
* 13960 14059: gap of 100 bp
* 14060 14785: contig of 726 bp in length
* 14786 14885: gap of 100 bp
* 14886 15580: contig of 695 bp in length
* 15581 15680: gap of 100 bp
* 15681 16420: contig of 740 bp in length
* 16421 16520: gap of 100 bp
* 16521 17233: contig of 713 bp in length
* 17234 17333: gap of 100 bp
* 17334 18057: contig of 724 bp in length
* 18058 18157: gap of 100 bp
* 18158 18874: contig of 717 bp in length
* 18875 18974: gap of 100 bp
* 18975 19716: contig of 742 bp in length
* 19717 19816: gap of 100 bp
* 19817 20557: contig of 741 bp in length
* 20558 20657: gap of 100 bp
* 20658 21379: contig of 722 bp in length
* 21380 21479: gap of 100 bp
* 21480 22193: contig of 714 bp in length
* 22194 22293: gap of 100 bp
* 22294 23013: contig of 720 bp in length
* 23014 23113: gap of 100 bp
* 23114 23851: contig of 738 bp in length
* 23852 23951: gap of 100 bp
* 23952 24681: contig of 730 bp in length
* 24682 24781: gap of 100 bp
* 24782 25505: contig of 724 bp in length
* 25506 25605: gap of 100 bp
* 25606 26335: contig of 730 bp in length
* 26336 26435: gap of 100 bp
* 26436 27158: contig of 723 bp in length
* 27159 27258: gap of 100 bp

```

```

* 27259 27960: contig of 702 bp in length
* 27961 28060: gap of 100 bp
* 28061 28775: contig of 715 bp in length
* 28776 28875: gap of 100 bp
* 28876 29586: contig of 711 bp in length
* 29587 29686: gap of 100 bp
* 29687 30394: contig of 708 bp in length
* 30395 30494: gap of 100 bp
* 30495 31246: contig of 752 bp in length
* 31247 31346: gap of 100 bp
* 31347 32090: contig of 744 bp in length
* 32091 32190: gap of 100 bp
* 32191 32894: contig of 704 bp in length
* 32895 32994: gap of 100 bp
* 32995 33741: contig of 747 bp in length
* 33742 33841: gap of 100 bp
* 33842 34602: contig of 761 bp in length
* 34603 34702: gap of 100 bp
* 34703 35440: contig of 738 bp in length
* 35441 35540: gap of 100 bp
* 35541 36261: contig of 721 bp in length
* 36262 36361: gap of 100 bp
* 36362 37081: contig of 720 bp in length
* 37082 37181: gap of 100 bp
* 37182 37898: contig of 717 bp in length
* 37899 37998: gap of 100 bp
* 37999 38730: contig of 732 bp in length
* 38731 38830: gap of 100 bp
* 38831 39526: contig of 696 bp in length
* 39527 39626: gap of 100 bp
* 39627 40316: contig of 690 bp in length
* 40317 40416: gap of 100 bp
* 40417 41136: contig of 720 bp in length
* 41137 41236: gap of 100 bp
* 41237 41985: contig of 749 bp in length
* 41986 42085: gap of 100 bp
* 42086 42815: contig of 730 bp in length
* 42816 42915: gap of 100 bp
* 42916 43632: contig of 717 bp in length
* 43633 43732: gap of 100 bp
* 43733 44445: contig of 713 bp in length
* 44446 44545: gap of 100 bp
* 44546 45287: contig of 742 bp in length
* 45288 45387: gap of 100 bp
* 45388 46129: contig of 742 bp in length
* 46130 46229: gap of 100 bp
* 46230 46961: contig of 732 bp in length
* 46962 47061: gap of 100 bp
* 47062 47765: contig of 704 bp in length
* 47766 47865: gap of 100 bp
* 47866 48582: contig of 717 bp in length
* 48583 48682: gap of 100 bp
* 48683 49429: contig of 747 bp in length
* 49430 49529: gap of 100 bp
* 49530 50236: contig of 707 bp in length
* 50237 50336: gap of 100 bp
* 50337 51062: contig of 726 bp in length
* 51063 51162: gap of 100 bp
* 51163 51884: contig of 722 bp in length
* 51885 51984: gap of 100 bp
* 51985 52702: contig of 718 bp in length
* 52703 52802: gap of 100 bp
* 52803 53516: contig of 714 bp in length
* 53517 53616: gap of 100 bp
* 53617 54329: contig of 713 bp in length
* 54330 54429: gap of 100 bp
* 54430 55140: contig of 711 bp in length
* 55141 55240: gap of 100 bp
* 55241 55930: contig of 690 bp in length
* 55931 56030: gap of 100 bp
* 56031 56724: contig of 694 bp in length
* 56725 56824: gap of 100 bp
* 56825 57628: contig of 804 bp in length

```

```

* 57629 57728: gap of 100 bp
* 57729 58523: contig of 795 bp in length
* 58524 58623: gap of 100 bp
* 58624 59387: contig of 764 bp in length
* 59388 59487: gap of 100 bp

Query Match          49.4%; Score 254.4; DB 63; Length 126330;
Best Local Similarity 66.4%; Pred. No. 2.8e-70;
Matches 314; Conservative 0; Mismatches 144; Indels 15; Gaps 2;

Qy 57 aggaagctgtgtacacagcatgataaaagacaatggagcgggtccacagtggtcccgctc 116
      ||| |||
Db 88902 ANNAACAAAGTATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 88843
Qy 117 ccttcagggtgatgagacgagctgtagagagatgtccagggaggttttcattaatca 176
      ||| |||
Db 88842 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 88783
Qy 177 gcaatttagtcagatctgtgcacatctcttatacaagaatgtcagtgggcc----- 229
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 88782 GCAATTTAGTCAAAATGCTACATCTCTATGCTTATTAAGAAATGTCAGTGGTCTTTCC 88723
Qy 230 -----tgagatcatcagatggaggttccatcggttttcaatgtcccgatccctttgt 282
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 88722 AAGGAGTGAGATCATCATGAAGTTCATTTGGTTTCAATGTCCTCGTATCCTTTGTA 88663
Qy 283 agacctgaagtggcaacgcaggaagaaacacagaaacccctggcggtgaattgcag 342
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 88662 AGACCTTGAAGTTGGCAATGTCAGGAAACAGGAACTCCACCTTAGCTCCATGAATTCG 88603
Qy 343 agctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 402
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 88602 AACTGTTGTGTTGTTTATGACCATCTGCCATCTTCTGTTTATGACACAGCTTGTGAA 88543
Qy 403 ctttaactggagctggggcaaaagtcaa-tcccacctttatacaaatgaattgtcgaagagg 461
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 88542 CTTTACTGAGAATGGTGAAGTAAATTCACAGTTTATACAAATGAATGCTGAAGAGG 88483
Qy 462 ccttttaaaactggagtgatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 514
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 88482 CCTTTAAAGTATAGAGTATGATGTTTATGGAAGTCTTTCCTATTAGGTC 88430

RESULT 7
AF327424/c
LOCUS      847 bp mRNA PLN 26-FEB-2001
DEFINITION Arabidopsis thaliana unknown protein (T14P1.19/At2g45010) mRNA,
            partial cds.
ACCESSION AF327424
VERSION AF327424.1 GI:11935196
KEYWORDS   Fli_CDNA.
SOURCE     thale cress.
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsids.
REFERENCE  1 (bases 1 to 847)
AUTHORS    Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
            Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L.,
            Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
            Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
            Satou,M., Seki,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T.,
            Davis,R.W., Ecker,J.R. and Theologis,A.
            Full length cDNA of gene T14P1.19/At2g45010 (GI:4895249)
            Unpublished
REFERENCE  2 (bases 1 to 847)
JOURNAL    Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
            Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L.,
            Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
            Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
            Jones,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T.,
            Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K.,

```



```
/gene="At2g44940"
2304. .>3809
/gene="At2g44940"
/feature="T14P1.26; pfam domain (PF00847)"
2322. .3809
/gene="At2g44940"
/codon_start=1
/product="putative AP2 domain transcription factor"
/db_xref="GI:4895256"
/translation="MARQINIESSVQVTFISSAIPAVSSSSITASASLSSTPTSS
SSSSSTNFIEDNRKASRSLSSLSVSDDDDDGGGRRKRTGDKPIPTG
VNRSGWKGIVIRPKRSRIWGLPTAEMARADHGQGGKRTGDKPIPTG
ELTPPTNSPKDQIAAASAAVNWQSDVNDVSEVAIEVAEPRAVVAQLFSSDTS
TTTTTQOSVSEASCASTACTDKSEEEKFLDPLDFTDENEMIRNDAPCYSSTW
QLCGDAGRLPEPFLSE"
complement(2972..3067)
/rpt_family="(GGA)n"
6558. .6661
/rpt_family="(TAAAA)n"
complement(6771..6870)
/rpt_family="(TAAAA)n"
complement(join(<8230..8361,8487..8652,8747..8949,
9035..9172,9411..9497,9590..9662,9774..9968,10342..10418,
10632..10703,10786..10946,11037..11255,11310..11343))
/gene="At2g44950"
complement(<8230..>11343)
/gene="At2g44950"
/notes="T14P1.25"
complement(join(8230..8361,8487..8652,8747..8949,
9035..9172,9411..9497,9590..9662,9774..9968,10342..10418,
10632..10703,10786..10946,11037..11255,11310..11343))
/gene="At2g44950"
/notes="unknown protein"
/codon_start=1
/protein_id="AAD32840.1"
/db_xref="GI:4895255"
/translation="MALLEKQVSVVVKLSIRKQVEKDSIVWKEINIKNELGV
SKTSVTSWASLEIQKQDERMIRKTRIGNTSRERGREIEFADMKALISFPE
EMSSRQNNYKETAGGHSRADVSGVLCRKEYAEALQSRADYASQGLDN
ATVDLKNHEELKFLDMV KRESTDARDIAEAKQKEYRAWHVQSLSLSDQNLLE
RYKANEAEVSOQMLAAAEATADLRQKMDCKRDVAKSDILKSKHEHGYLSIE
QTIGSYEDIVFQNGQLLQVTFEDDYNINIKFLLEGTSRQMDTLIDKIMKDIQI
GAYASFLSKSSRIEDQLRFCTQFKLAEDRYQSVLENQKLRSLRGNGEQAR
SKLESHSVESGSLDGALELEIRFNRRRIEEMETAKKKVRSLSRSLIGSSAI
QKLRQSLSEFKTELKACANDRPKEVITKCYHLFCNCPVQKLTGTQRKKCPTCSASF
GPNDKPIVI"
complement(join(<12607..12809,12908..13135,13255..13336,
13417..13578,13759..>13875))
/gene="At2g44960"
complement(<12607..>13875)
/notes="T14P1.24; similar to hypothetical protein
GB: AAC31828"
complement(join(12607..12809,12908..13135,13255..13336,
13417..13578,13759..13875))
/gene="At2g44960"
/notes="hypothetical protein"
/codon_start=1
/protein_id="AAD32839.1"
/db_xref="GI:4895254"
/translation="MASTGEPRKRRHFSISPSERAAAVKKQFFWPPSSDKLDTAV
LQFNILSKLEAQVECSILEDKLSQKEKQLPYNSSILKTVHKSWEKQITASVESC
SVYSDSSGSAFVNKPGSSPAVKNDFINRLLETGATSSSNITCSNOMRNGVNT
SSMOTQTLNVAATDELCKLDKDELYPTVLRNLGKDLGQALSELSELETSFPRGDL
DVLVFKLSRLSELQSHRDADAKVRVDRKLRGIRSYLVSSHSTIYC"
14535. 14617
/rpt_family="(TAAAA)n"
complement(join(<14627..15100,15283..15570,15651..15738,
15995..16123,16289..16535,16728..16836,17072..>17122))
/gene="At2g44970"
complement(<14627..>17122)
/gene="At2g44970"

/note="T14P1.23; identical to GB: AAC31827"
complement(join(14627..15100,15283..15570,15651..15738,
15995..16123,16289..16535,16728..16836,17072..17122))
/gene="At2g44970"
/notes="unknown protein"
/codon_start=1
/protein_id="AAD32838.1"
/db_xref="GI:4895253"
/translation="MGEAGSDGSEVPLVGGGNDNAVAPQVFNLSLPALNEAASYI
TOATSYLGSLEYGKDTISHPHELLRSTSGVDGNSPVSVCSISGPRPSSSASTDA
NSPRESTEIIPQATNAIVTSNRLNLGISMFOGLIERARRVRGASDDIGLQRAPE
MPVEDGDRFNKILELIGNHGPLYFDTTKTKSKMGKACHIAKHSSESVKNAEII
KEYTEELGNSKRIKLLGHGSGGIDAAALSYWPELKDVLGVLKHSVSGKGSRTIA
TDIREGOLGYVNLKRMKEILISKVIGKIDQALDELTYRERKEFKLGNPLRLEPTV
SFRTASTSPAVLSTLSHVAHAELPLTQAAKLPVMPGLGAAACAOQLQVRYGKGS
DGLVTCDDAEVPGSVVVRKRLDHAWVSYSSNEVPLEADAAQVCEALLTLVQVVEQ
ERQKQKATKID"
complement(join(<18068..18186,18313..18477,18558..18703,
18775..19400,19477..19602,19724..19924,20006..20113,
20198..20287,20397..20514,20614..20749,20875..21046,
21269..21400,21551..21684,21782..21871,22075..>22297))
/gene="At2g44980"
complement(<18068..>22297)
/gene="At2g44980"
/notes="T14P1.22"
complement(join(18068..18186,18313..18477,18558..18703,
18775..19400,19477..19602,19724..19924,20006..20113,
20198..20287,20397..20514,20614..20749,20875..21046,
21269..21400,21551..21684,21782..21871,22075..22297))
/gene="At2g44980"
/codon_start=1
/product="SNF2 subfamily global transcription activator"
/protein_id="AAD32837.1"
/db_xref="GI:4895252"
/translation="MSKESPPKVPSTTMEYRRLEAAAEIILEKAEKFSNTPDCSE
EGVATATLPHQVEGVSLIQKLVLLGVNVLELDOMGLCTLOALSEFLSKRQGLPG
PFLVCLPSVTDGWSVSEINRFTPNLEVLRYGDKYCRDMRKSMDHGFHFLFDVLIT
TYDALVDQDFSLQIPQWYAIIDEAQRKNPNVLYNLEQLFLPRLLITGTPTON
NLTEALWHFCMPLVFTLODFSAFKETGDLGVNDKETYKSLKFLIGAFMLRRT
KSLITESCNIPLVPLTELTMVPLVSLQAKIYTSILRKELGELLESGGNSHTSQN
IVIOLRKASHPYLPFGTPEPFEHGLVQASGLLVLDQLLRHDSGHRVLLFSQ
MSTLIDILQDMELRYSYERLDGSVRAEERFAALKNFSAKTERGLDSEVDSNAEVE
MISPRAGVGLNLVADRVIFYEQDWNQVQKQALQRAHRIQISHVLSINLVTESHV
EEVILRAERKLQSLHNVDNNEKEEDGDLRLVFGIQRFPDPEIHNESDNLKM
VEISLAQVVAIRQNVDPKREINSDTLIGNTSSALSDPEIHNESDNLKM
LKEARSKDEKIIILGNKRLSEERNLIEAARKKAEKKLATWGAGHYQSLSVEEP
```


TITLE
JOURNAL
REFERENCE
AUTHORS

Homo sapiens chromosome 4, clone RP11-70M19
Unpublished
2 (bases 1 to 162299)
Birren, S., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bada, F., Boguslavsky, L.,
Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Ferreira, P.,
FitzHugh, W., Gage, D., Galaan, J., Gardyna, S., Ginde, S., Goyette, M.,
Graham, L., Grand-pierre, N., Hagos, B., Heaford, A., Horton, L.,
Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., LaRocque, K.,
Lamarez, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G.,
Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McKernan, K.,
McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,
Morrow, J., Murphy, T., Taylor, J., Norman, C. H., O'Connor, T.,
O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K.,
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Rieback, M., Riley, R.,
Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,
Sounez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,
Zimmer, A. and Zody, M.
Direct Submission
Submitted (18-AUG-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 18, 2000 this sequence version replaced gi:9845128.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center

TITLE
JOURNAL
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L10622
Center clone name: 70_M_19
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 149171 bases at least Q40
Consensus quality: 155807 bases at least Q30
Consensus quality: 158244 bases at least Q20
Insert size: 163000; agarose-fp
Insert size: 159499; sum-of-contigs
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 3.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

```

1 1088: contig of 1088 bp in length
* 1089 1188: gap of 100 bp
* 1189 2239: contig of 1051 bp in length
* 2240 2339: gap of 100 bp
* 2340 3811: contig of 1472 bp in length
* 3812 3911: gap of 100 bp
* 3912 5158: contig of 1247 bp in length
* 5159 5258: gap of 100 bp
* 5259 6908: contig of 1650 bp in length
* 6909 7008: gap of 100 bp
* 7009 8568: contig of 1560 bp in length
* 8569 8668: gap of 100 bp
* 8669 10761: contig of 2093 bp in length
* 10762 10861: gap of 100 bp
* 10862 13026: contig of 2165 bp in length

```

```

* 13027 13126: gap of 100 bp
* 13127 15329: contig of 2203 bp in length
* 15330 15429: gap of 100 bp
* 15430 17434: contig of 2005 bp in length
* 17435 17534: gap of 100 bp
* 17535 19935: contig of 2401 bp in length
* 19936 20035: gap of 100 bp
* 20036 23899: contig of 3864 bp in length
* 23900 23999: gap of 100 bp
* 24000 27393: contig of 3394 bp in length
* 27394 27493: gap of 100 bp
* 27494 47629: contig of 20136 bp in length
* 47630 47729: gap of 100 bp
* 47730 51364: contig of 3635 bp in length
* 51365 51464: gap of 100 bp
* 51465 55977: contig of 4513 bp in length
* 55978 56077: gap of 100 bp
* 56078 60357: contig of 4280 bp in length
* 60358 60457: gap of 100 bp
* 60458 64814: contig of 4357 bp in length
* 64815 64914: gap of 100 bp
* 64915 70287: contig of 5373 bp in length
* 70288 70387: gap of 100 bp
* 70388 75397: contig of 5010 bp in length
* 75398 75497: gap of 100 bp
* 75498 81554: contig of 6057 bp in length
* 81555 81654: gap of 100 bp
* 81655 88042: contig of 6388 bp in length
* 88043 88142: gap of 100 bp
* 88143 95604: contig of 7462 bp in length
* 95605 95704: gap of 100 bp
* 95705 102851: contig of 7147 bp in length
* 102852 102951: gap of 100 bp
* 102952 114380: contig of 11429 bp in length
* 114381 114480: gap of 100 bp
* 114481 126822: contig of 12342 bp in length
* 126823 126922: gap of 100 bp
* 126923 141300: contig of 14378 bp in length
* 141301 141400: gap of 100 bp
* 141401 159634: contig of 18234 bp in length
* 159635 159734: gap of 100 bp
* 159735 162299: contig of 2565 bp in length.

```

FEATURES
Source
1..162299
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone_lib="RP11-70M19"
/clone="RP11-70M19" Human Male BAC"
1..1088
/note="assembly_fragment
clone_end:SP6
vector_side:left"
1189..2239
/note="assembly_fragment"
2340..3811
/note="assembly_fragment"
3912..5158
/note="assembly_fragment"
5259..6908
/note="assembly_fragment"
7009..8568
/note="assembly_fragment"
8669..10761
/note="assembly_fragment"
10862..13026
/note="assembly_fragment"
13127..15329
/note="assembly_fragment"
15430..17434
/note="assembly_fragment"
17535..19935

misc_feature
1..1088
/note="assembly_fragment
clone_end:SP6
vector_side:left"
1189..2239
/note="assembly_fragment"
2340..3811
/note="assembly_fragment"
3912..5158
/note="assembly_fragment"
5259..6908
/note="assembly_fragment"
7009..8568
/note="assembly_fragment"
8669..10761
/note="assembly_fragment"
10862..13026
/note="assembly_fragment"
13127..15329
/note="assembly_fragment"
15430..17434
/note="assembly_fragment"
17535..19935

2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsukuba@affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps is believed
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

```

FEATURES
source
  Location/Qualifiers
    1. .157721
      /organism="Oryza sativa"
      /cultivar="Nipponbare"
      /db_xref="taxon:4530"
      /chromosome="1"
      /clone="P0490D09"
BASE COUNT  44929 a 33979 c 34269 g 44344 t 200 others
ORIGIN
Query Match          7.2%; Score 37; DB 83; Length 157721;
Best Local Similarity 56.0%; Pred. No. 2.3;
Matches 70; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY  197  cactcatgctttacaagaatgtcagtgggccctgagatcatcatcagatggaggttcacgg 256
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   24115 CATCCAATCTTTTAAAGAGACTCCAGAGGTTCTGGAAGGCCAGATCAATCTTCAGTAA 24174

QY  257  gtttcaatgccggtatccctttgttaagacctgaagttagcagcagagaaaaacaggaa 316
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   24175 CTATCTTGATGCTGATATTAAATGCATCACCTTCAAAAGTTTGGTGCCAGTAAGAAGTCA 24234

QY  317  ctcca 321
      |||||
Db  24235 CTCCA 24239

```

RESULT	14
AF217246/c	
LOCUS	AF217246 182615 bp DNA HTG 05-SEP-2000
DEFINITION	Homo sapiens chromosome 8 clone CTD-2017M2 map 8q24, WORKING DRAFT SEQUENCE, 21 unordered pieces.
ACCESSION	AF217246
VERSION	AF217246.4 GI:9967152
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	Taudien,S., Blechschmidt,K., Menzel,U., Polley,A., Reichwald,K., Rump,A., Schilhabel,M.B., Schudy,A., Wen,G. and Rosenthal,A. Chromosome 8 genomic sequence Unpublished 2 (bases 1 to 182615) Taudien,S., Wen,G.P., Schilhabel,M., Menzel,U., Jahn,N., Baumgart,C. and Rosenthal,A. Direct Submission Submitted (16-DEC-1999) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
AUTHORS	3 (bases 1 to 182615) Taudien,S., Blechschmidt,K., Menzel,U., Polley,A., Reichwald,K., Rump,A., Schilhabel,M., Schudy,A., Wen,G. and Rosenthal,A. Direct Submission Submitted (05-SEP-2000) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
JOURNAL	Sequence update by submitter On Sep 5, 2000 this sequence version replaced gi:8152189. ★ NOTE: This is a 'working draft' sequence. It currently
REMARK	
COMMENT	

* consists of 21 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

1	9231:	contig of 9231 bp in length
9232	9331:	gap of unknown length
9332	10427:	contig of 1096 bp in length
10428	10527:	gap of unknown length
10528	12883:	contig of 2356 bp in length
12884	12983:	gap of unknown length
12984	14772:	contig of 1789 bp in length
14773	14872:	gap of unknown length
14873	18731:	contig of 3859 bp in length
18732	18831:	gap of unknown length
18832	22448:	contig of 3617 bp in length
22449	22548:	gap of unknown length
22549	26063:	contig of 3515 bp in length
26064	26163:	gap of unknown length
26164	30412:	contig of 4249 bp in length
30413	30512:	gap of unknown length
30513	37855:	contig of 7343 bp in length
37856	37955:	gap of unknown length
37956	48596:	contig of 10641 bp in length
48597	48696:	gap of unknown length
48697	58215:	contig of 9519 bp in length
58216	58315:	gap of unknown length
58316	72799:	contig of 14484 bp in length
72800	72899:	gap of unknown length
72900	86382:	contig of 13483 bp in length
86383	86482:	gap of unknown length
86483	100353:	contig of 13871 bp in length
100354	100453:	gap of unknown length
100454	113432:	contig of 12979 bp in length
113433	113532:	gap of unknown length
113533	130872:	contig of 17340 bp in length
130873	130972:	gap of unknown length
130973	156022:	contig of 25050 bp in length
156023	156122:	gap of unknown length
156123	178374:	contig of 22252 bp in length
178375	178474:	gap of unknown length
178475	179488:	contig of 1014 bp in length
179489	179588:	gap of unknown length
179589	181002:	contig of 1414 bp in length
181003	181102:	gap of unknown length
181103	182615:	contig of 1513 bp in length

FEATURES	source
	misc_f
	.
BASE COUNT	
ORIGIN	
Query Match	
Best Local	
Matches	
Qy	337 t
	t
Db	128578 T
	T
Qy	397 t
	t

	Query Match	7.2%	Score 37	DB 78	Length 182615	
	Best Local Similarity	56.0%	Pred. No. 2.3			
	Matches	70	Conservative	0	Mismatches	55
					Indels	0
					Gaps	0
QY	337	ttgcagagcgtttdgtgtgtttgtgaccactctcccattcttctctgttatgacagagct	396			
Db	128578	TTGTTGAGTTTTATGTGTGTTCTGTGAACTTCAGATACTATGTTTTTCCCTTAGAG	128519			
QY	397	tgtgaaccttaactgggagctggggcgaagtccaatcccacacctttacacatgaattgctga	456			

Db 128518 TTGCTATTCAAGATGGGGTGAGAAAAAGAACAGCTGACCCCTTGAAACATACAGGTTAGA 128459

QY 457 agagg 461

Db 128458 ACTGG 128454

RESULT 15

AL390758/c

LOCUS

DEFINITION

Homo sapiens chromosome 10 clone RP11-540C21, *** SEQUENCING IN

PROGRESS ***

ACCESSION

AL390758

VERSION

AL390758.3

KEYWORDS

SOURCE

HTG; HTGS_PHASE1; HTGS_DRAFT.

human

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

SimS, S.

Direct Submission

Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SP, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Aug 25, 2000 this sequence version replaced gi:9716790.

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: BA540C21

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 136465 bases at least Q40

Consensus quality: 141338 bases at least Q30

Consensus quality: 144184 bases at least Q20

Insert size: 146267; sum-of-contigs

Insert size: 166241; 2.3% error; agarose-fp

Quality coverage: 3.25x in Q20 bases; sum-of-contigs Quality

coverage: 3.06x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently

* consists of 21 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 4224: contig of 4224 bp in length

* 4225 4324: gap of 100 bp

* 4325 10808: contig of 6484 bp in length

* 10809 10908: gap of 100 bp

* 10909 26777: contig of 15869 bp in length

* 26778 26877: gap of 100 bp

* 26878 31189: contig of 4312 bp in length

* 31190 31289: gap of 100 bp

* 31290 40015: contig of 8726 bp in length

* 40016 40115: gap of 100 bp

* 40116 52786: contig of 12671 bp in length

* 52787 52886: gap of 100 bp

* 52887 55454: contig of 2568 bp in length

* 55455 55554: gap of 100 bp

* 55555 59896: contig of 4342 bp in length

* 59897 59996: gap of 100 bp

* 59997 64419: contig of 4423 bp in length

* 64420 64519: gap of 100 bp

* 64520 66621: contig of 2102 bp in length

* 66622 66721: gap of 100 bp

* 66722 70075: contig of 3354 bp in length

* 70076 70175: gap of 100 bp

* 70176 88176: contig of 18001 bp in length

* 88177 88276: gap of 100 bp

* 88277 90769: contig of 2493 bp in length

* 90770 90869: gap of 100 bp

* 90870 96615: contig of 5746 bp in length

* 96616 96715: gap of 100 bp

* 96716 101018: contig of 4303 bp in length

* 101019 101118: gap of 100 bp

* 101119 106502: contig of 5384 bp in length

* 106503 106602: gap of 100 bp

* 106603 109577: contig of 2975 bp in length

* 109578 109677: gap of 100 bp

* 109678 129521: contig of 19844 bp in length

* 129522 129621: gap of 100 bp

* 129622 134344: contig of 4723 bp in length

* 134345 134444: gap of 100 bp

* 134445 142809: contig of 8365 bp in length

* 142810 142909: gap of 100 bp

* 142910 148267: contig of 5358 bp in length.

FEATURES

source

1. 148267

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="10"

/clone="RP11-540C21"

/clone_lib="RPC1-11.2"

1. 4224

/note="assembly_fragment:00232"

fragment_chain:1

4325. 10808

/note="assembly_fragment:00914"

fragment_chain:1

10909. 26777

/note="assembly_fragment:00153"

26878. 31189

/note="assembly_fragment:00185"

31290. 40015

/note="assembly_fragment:00235"

40116. 52786

/note="assembly_fragment:00323"

52887. 55454

/note="assembly_fragment:00363"

55555. 59896

/note="assembly_fragment:00401"

59997. 64419

/note="assembly_fragment:00466"

64520. 66621

/note="assembly_fragment:00473"

66722. 70075

/note="assembly_fragment:00517"

70176. 88176

/note="assembly_fragment:00531"

88277. 90769

/note="assembly_fragment:00542"

90870. 96615

/note="assembly_fragment:00568"

96716. 101018

/note="assembly_fragment:00571"

101119. 106502

/note="assembly_fragment:00784"

106603. 109577

/note="assembly_fragment:00869"

109678. 129521

/note="assembly_fragment:00907"

129622. 134344

/note="assembly_fragment:00990"

134445. 142809

/note="assembly_fragment:01164"

142910. 148267

/note="assembly_fragment:00253"

clone_end:SP6

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 18:00:40 ; Search time 346.98 Seconds
(without alignments)
866.469 Million cell updates/sec

Title: US-09-445-201-1-copy_10094_10608

Perfect score: 515

Sequence: 1 tttaaatgtgctgtcttag.....agggttccttcttgatcc 515

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0401.*
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	515	100.0	12845	20 V84274	Mouse endothelial
C 2	39.4	7.7	1044	21 C45640	Arabidopsis thalia
C 3	39.4	7.7	1046	21 C33819	Arabidopsis thalia
C 4	33.8	6.6	2274	20 X84333	Stealth virus nucl
5	32.2	6.3	92334	21 A81473	N. meningitidis pa
6	32.2	6.3	349980	21 F21544	Neisseria meningit
7	32.2	6.3	1437668	21 A81490	N. meningitidis B
C 8	31.6	6.1	5520	21 C51145	Arabidopsis thalia
9	30.8	6.0	1222	20 X37475	Human secreted pro
10	30.8	6.0	5612	21 C76389	Human ORFX ORF1944
11	30.6	5.9	314	19 V31227	E. coli J96 pathog

12	30.6	5.9	11165	21 A15186	DNA encoding Esche
13	30.6	5.9	273254	21 C81914	Chlamydia pneumoni
14	30.4	5.9	3161	21 C74253	Human secreted pro
C 15	30	5.8	692	20 Z42012	Human endometrium
16	30	5.8	7679	21 Z36322	Mechanical stress
17	30	5.8	8882	21 Z36321	Mechanical stress
C 18	29.8	5.8	550	21 C52708	Arabidopsis thalia
C 19	29.8	5.8	1590	21 Z99655	DNA coding for the
20	29.8	5.8	2105	21 C59480	Human secreted pro
21	29.8	5.8	2432	21 A79269	Pinus radiata cell
C 22	29.6	5.7	308	21 C16102	Human secreted pro
23	29.6	5.7	568	21 X00667	Human colon cancer
C 24	29.6	5.7	2693	20 X02976	Human IL-1ra BAC c
25	29.4	5.7	704	22 F44760	Annexin-like prote
26	29.4	5.7	2429	21 A98003	Human T gene DNA f
27	29.2	5.7	648	21 F10383	Fusarium venenatum
28	29	5.6	397	8 N71366	Sequence encoding
C 29	29	5.6	837	21 Z36301	Mechanical stress
30	29	5.6	886	12 Q10077	Engineered human T
31	29	5.6	886	17 T41697	TSH beta subunit g
C 32	29	5.6	1947	20 X52247	Protein PRQ239 cDN
C 33	29	5.6	2384	19 V30919	Human secreted pro
C 34	29	5.6	2641	21 Z54227	HTM clone 2709055
C 35	29	5.6	5077	21 A50096	Arabidopsis herbic
36	28.8	5.6	535	21 C54997	Arabidopsis thalia
37	28.8	5.6	704	21 A67096	Pinus radiata anne
38	28.8	5.6	704	21 A67132	Pinus radiata anne
C 39	28.8	5.6	1420	20 X90222	GRK4 polymorphism
40	28.8	5.6	1815	21 Z54283	Neisseria gonorrhe
41	28.8	5.6	2088	18 T96687	Human TUB form 6 c
42	28.8	5.6	2088	21 A94680	Human TUB form 6 c
43	28.6	5.6	440	21 F15627	Human prostate can
44	28.6	5.6	716	22 C90339	AI378857 cDNA clon
C 45	28.6	5.6	2346	21 A95821	Human metalloprote

ALIGNMENTS

RESULT 1	
ID V84274	V84274 standard; DNA; 12845 BP.
XX AC V84274;	
XX DT 12-APR-1999 (first entry)	
XX DE Mouse endothelial growth factor receptor-2 gene Flk-1 5' region.	
XX KW Flk-1; endothelial growth factor receptor-2; VEGF; transcription;	
KW endothelium; enhancer; vascular disease; angiogenesis; cancer;	
KW diabetic retinopathy; rheumatoid arthritis; wound healing;	
KW vulneryary; atherosclerosis; tumour; neuronal disorder; therapy;	
KW diagnosis; mouse; ss.	
XX OS Mus sp.	
XX FH Key	Location/Qualifiers
FT promoter	6036..6959
FT exon	/tag= a
FT	6661..7026
FT	/tag= b
FT	/number= 1
FT	/codon_start= 6960
FT	7027..10632
FT	/tag= c
FT	/number= 1
FT	/note= "DNA sequences comprising nucleotides
FT	8260-10560, 8336-10608 and/or 10094-10608
FT	are specifically claimed in Claim 3b"
FT	protein_bind 10137..10155
FT	/tag= d
FT	/note= "GATA/PEA3 consensus binding site"

```

FT protein_bind 10166..10179
FT /tag= e
FT /note= "GATA consensus binding site"
FT protein_bind 10187..10197
FT /tag= f
FT protein_bind 10252..10261
FT /note= "AP1 consensus binding site"
FT protein_bind 10273..10283
FT /tag= g
FT /note= "NFkB consensus binding site"
FT protein_bind 10301..10309
FT /tag= h
FT /note= "AP1 consensus binding site"
FT protein_bind 10356..10372
FT /tag= i
FT /note= "STAT consensus binding site"
FT protein_bind 10395..10404
FT /tag= j
FT /note= "Ets-1/GATA consensus binding site"
FT protein_bind 10453..10477
FT /tag= k
FT /note= "Ets-1 consensus binding site"
FT protein_bind 10643..10726
FT /tag= l
FT /note= "SCL/TAL-1 Ets-1 consensus binding site"
FT exon 10727..12673
FT /number= 2
FT intron 12674..12845
FT /number= 2
FT exon 12846..12845
FT /tag= o
FT /number= 3
FT PN W09855638-Al.
FT PD 10-DEC-1998.
FT XX 03-JUN-1998; 98WO-EF03318.
FT PR 03-JUN-1997; 97EP-0108959.
FT XX (PLAC ) MAX PLANCK GRS FOERDERUNG WISSENSCHAFTEN.
FT PA Breier G, Risau W, Roenicke V;
FT PI WPI; 1999-059915/05.
FT XX
FT PT New recombinant DNA containing heterologous sequence controlled by
FT endothelial cell specific regulator - from the Flk-1 gene, used to
FT treat, prevent or diagnose vascular disease, tumours, also to screen
FT for transcriptional regulators
FT PS Claim 3; Fig 1; 107pp; English.
FT XX
CC This 12.8 kb DNA sequence spans the region from -6.65 kb relative to
CC the transcriptional start site to +6.15 (located in the third exon)
CC of the murine endothelial growth factor (VEGF) receptor-2 gene
CC Flk-1. This portion of the Flk-1 gene was isolated from the liver
CC of mouse strain 129/SvJ by screening a phage library of liver
CC genomic DNA in vector lambda FixII. Sequences within the 5'-flanking
CC region of the gene, in combination with sequences located within the
CC first intron, specifically and reproducibly target expression of
CC heterologous DNA to angioblasts during early stages of vascular
CC development and also to the vasculature of postnatal mice. The
CC regulatory sequences of the first intron also function as an
CC autonomous endothelium-specific enhancer when fused to a
CC heterologous promoter (e.g. HSV-TK promoter). This Flk-1 intron
CC enhancer contains several potential binding sites for transcription
CC factors of the Ets and GATA families. The invention provides a
CC new claimed recombinant DNA (I) comprising at least one regulatory
CC sequence from an intron of the Flk-1 gene, or its homologue, to
CC control expression in endothelial cells, in vivo, linked to (b) a

```

```

CC heterologous DNA. (I) is used to produce transgenic animals, and
CC these, or transformed cells, are used to identify agents (A),
CC potential pharmaceuticals, that suppress, activate or enhance
CC transcription of genes in endothelial cells. (I), vectors and (A)
CC are used to direct or prevent (for antisense sequences) expression
CC of genes specifically in endothelial cells, e.g. for treating
CC angiogenesis, cancer, diabetic retinopathy, rheumatoid arthritis
CC etc., and in wound healing, particularly for treating vascular
CC disease (particularly proliferation of smooth muscle cells,
CC specifically atherosclerosis), tumours and neuronal disorders.
CC They may also be used to induce vascular disease in the transgenic
CC animals or diagnostically, particularly for studying (mal)function,
CC interactions and unregulated expression of endothelial products.
CC (I) provide specific modulation of gene expression in endothelial
CC cells, at all stages of development.
XX
SQ Sequence 12845 BP; 3321 A; 2893 C; 2863 G; 3723 T; 45 other;

Query Match 100.0%; Score 515; DB 20; Length 12845;
Best Local Similarity 100.0%; Pred. No. 9.7e-164;
Matches 515; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttttaaatgtgtcttttagaagccactgcctcagcttctcagctcagatcccaaggga 60
Db 10094 ttttaaatgtgtcttttagaagccactgcctcagcttctcagctcagatcccaaggga 10153
Qy 61 agtctggtacacagcatgataaaagacaaatggagcgggtcacagtggctcccgctccctt 120
Db 10154 agtctggtacacagcatgataaaagacaaatggagcgggtcacagtggctcccgctccctt 10213
Qy 121 tcagggggtatggagacgagctgtagagagatgtctccaggagggttttcattaatcagcaa 180
Db 10214 tcagggggtatggagacgagctgtagagagatgtctccaggagggttttcattaatcagcaa 10273
Qy 181 tttagtcagatctgtgcattcctctatctttacaagaataatcagtgaggcctgagatcatca 240
Db 10274 tttagtcagatctgtgcattcctctatctttacaagaataatcagtgaggcctgagatcatca 10333
Qy 241 gatggaggttcacgtgggtttcaatgtcccgatcccttttgaagaccttgaagtggcaa 300
Db 10334 gatggaggttcacgtgggtttcaatgtcccgatcccttttgaagaccttgaagtggcaa 10393
Qy 301 cgcaggaaaacaggaactccaccctcgtggtcgtgaattgcagagctgtgtgtgtttg 360
Db 10394 cgcaggaaaacaggaactccaccctcgtggtcgtgaattgcagagctgtgtgtgtttg 10453
Qy 361 tgaccatctgccattcttcctgttatgacagagcttggaactttaactggagctgggg 420
Db 10454 tgaccatctgccattcttcctgttatgacagagcttggaactttaactggagctgggg 10513
Qy 421 caaagtcacatccaccctttatacaatgaattgctgaagagggccttttaaaacttgagtg 480
Db 10514 caaagtcacatccaccctttatacaatgaattgctgaagagggccttttaaaacttgagtg 10573

RESULT 2
C45640/c
ID C45640 standard; DNA; 1044 BP.
XX
AC C45640;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 47226.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.

```

XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX XD
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.

Query Match	Score 39.4;	DB 21;	Length 1044;
Best Local Similarity	71.2%;	Pred. No. 0.0037;	
Matches 52;	Conservative	0;	Mismatches 21; Indels 0; Gaps 0;
Pr 13-SEP-1999;	99US-01537558.		
Pr 15-SEP-1999;	99US-0154018.		
Pr 16-SEP-1999;	99US-0154039.		
Pr 20-SEP-1999;	99US-0154779.		
Pr 22-SEP-1999;	99US-0155139.		
Pr 23-SEP-1999;	99US-0155486.		
Pr 24-SEP-1999;	99US-0155659.		
Pr 28-SEP-1999;	99US-0156458.		
Pr 29-SEP-1999;	99US-0156596.		
Pr 04-OCT-1999;	99US-0157117.		
Pr 05-OCT-1999;	99US-0157753.		
Pr 06-OCT-1999;	99US-0157865.		
Pr 07-OCT-1999;	99US-0158029.		
Pr 08-OCT-1999;	99US-0158232.		
Pr 12-OCT-1999;	99US-0158369.		
Pr 13-OCT-1999;	99US-0159293.		
Pr 13-OCT-1999;	99US-0159294.		
Pr 13-OCT-1999;	99US-0159295.		
Pr 14-OCT-1999;	99US-0159329.		
Pr 14-OCT-1999;	99US-0159330.		
Pr 14-OCT-1999;	99US-0159331.		
Pr 14-OCT-1999;	99US-0159637.		
Pr 14-OCT-1999;	99US-0159638.		
Pr 18-OCT-1999;	99US-0159584.		
Pr 21-OCT-1999;	99US-0160741.		
Pr 21-OCT-1999;	99US-0160767.		
Pr 21-OCT-1999;	99US-0160770.		
Pr 21-OCT-1999;	99US-0160814.		
Pr 21-OCT-1999;	99US-0160815.		
Pr 22-OCT-1999;	99US-0160980.		
Pr 22-OCT-1999;	99US-0160981.		
Pr 22-OCT-1999;	99US-0160989.		
Pr 25-OCT-1999;	99US-0161404.		
Pr 25-OCT-1999;	99US-0161405.		
Pr 25-OCT-1999;	99US-0161406.		
Pr 26-OCT-1999;	99US-0161359.		
Pr 26-OCT-1999;	99US-0161360.		
Pr 26-OCT-1999;	99US-0161361.		
Pr 28-OCT-1999;	99US-0161920.		
Pr 28-OCT-1999;	99US-0161992.		
Pr 28-OCT-1999;	99US-0161993.		
Pr 29-OCT-1999;	99US-0162142.		
Pr 187	cagatctgtgcattcctatgctttcaagaagaatgtcagtgaggcctgagatcatcatcgatgga	246	
Db	360 CAGCTCTCTGGATCTCAGCACACCAAAAATGCGGGTGTCCAGTTTTCATCAGAAGGA	301	
Qy	247 ggttcacggggtt	259	
Db	300 GGTTCATAGGTTT	288	
RESULT 3			
ID C33819/c			
XX C33819.			
XX C33819.			
DT 17-OCT-2000	(first entry)		
XX Arabidopsis thaliana	DNA fragment SPQ ID NO: 4426.		
XX Hybridisation assay;	genetic mapping; gene expression control;		
KW protein identification;	signal transduction pathway;		
KW metabolic pathway;	promoter; termination sequence; ss.		
XX Arabidopsis thaliana.			
OS Arabidopsis thaliana.			
XX Arabidopsis thaliana.			

PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146388.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 10-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 23-AUG-1999; 99US-0151080.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151330.
 PR 01-SEP-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0152363.
 PR 13-SEP-1999; 99US-0153070.
 PR 15-SEP-1999; 99US-0153758.
 PR 16-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 22-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159294.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.
 PR 29-OCT-1999; 99US-0162142.
 Query Match 6.1%; Score 31.6; DB 21; Length 5520;
 Best Local Similarity 55.5%; Pred. No. 3.9;
 Matches 61; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
 QY 67 gtacacagcatgataaaagacaaatggagcgggtgcacagtggtcccgctcccttcagg 126
 Db 3739 GCACATAAGATGAAAGAAAGGTGAGAGTGGTCACTTGTCTAATGCCCCGTTCTGGT 3680
 QY 127 gtatggagcagcgtgtagagagatgtctccagggaggttttcattatca 176
 Db 3679 TGTCAAACCAAGTCGAGAGAAAGCTTCAAGGAAGTTCCATTCAACA 3630
 RESULT 9
 X37475
 ID X37475 standard; cDNA; 1222 BP.
 XX
 AC X37475;
 XX
 DT 06-JUL-1999 (first entry)
 XX
 DE Human secreted protein cDNA fragment containing gene 25.
 XX
 KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
 KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
 KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
 KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
 KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
 KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
 KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
 KW arthritis; malignancy; digestive; endocrine; infection; ss.
 XX Homo sapiens.
 OS
 XX WO9918208-A1.
 PN
 PD 15-APR-1999.
 XX
 PF 01-OCT-1998; 98WO-US20775.
 XX
 PR 02-OCT-1997; 97US-0060884.
 PR 02-OCT-1997; 97US-0060833.
 PR 02-OCT-1997; 97US-0060836.
 PR 02-OCT-1997; 97US-0060837.
 PR 02-OCT-1997; 97US-0060838.
 PR 02-OCT-1997; 97US-0060839.
 PR 02-OCT-1997; 97US-0060843.
 PR 02-OCT-1997; 97US-0060862.
 PR 02-OCT-1997; 97US-0060866.
 PR 02-OCT-1997; 97US-0060874.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Carter KC, Duan DR, Endress GA, Feng P, Ferlie AM;
 PI Florence KA, Greene JM, Janat F, Lafleur DW, Ni J;
 PI Rosen CA, Ruben SM, Shi Y, Young P, Yu G;
 XX WPI; 1999-264022/22.
 DR P-PSDB; Y07876.
 XX
 PT New isolated human genes and the secreted polypeptides they encode
 PS Claim 1a; Page 241-242; 368pp; English.
 XX This invention describes novel isolated human genes and the secreted


```

PN WO9822575-A2.
XX
PD 28-MAY-1998.
XX
PF 21-NOV-1997; 97WO-US21347.
XX
PR 14-OCT-1997; 97US-0061953.
PR 22-NOV-1996; 96US-0031626.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (UYWI-) UNIV WISCONSIN.
XX
PI Choi GH, Dillon PJ, Welch RA;
XX WPI; 1998-312461/27.
DR
XX
XX New isolated uropathogenic E. coli nucleotide sequences - used to
PT develop products for the detection of pathogenic E. coli and to
PT elicit an immune response to pathogenic E. coli
XX
XX Claim 21; Page 139-140; 250pp; English.
XX
XX This sequence represents a E. coli strain J96 contig containing
CC pathogenicity island (PAI) sequences, and represents a nucleic acid
CC molecule of the invention. PAIs are large fragments of DNA which comprise
CC pathogenicity determinants. The sequences of the invention are taken from
CC PAI IV and PAI V. PAI IV is located at approximately 64 min (near pheV)
CC on the E. coli chromosome and is greater than 170 kb. PAI V is located at
CC approximately 94 min (at pheR) on the E. coli chromosome and is
CC approximately 160 kb in size. Antibodies specific to the proteins encoded
CC by the PAI open reading frames of the invention can be used in kits to
CC detect uropathogenic E. coli. The proteins are used in vaccines to elicit
CC a protective immune response in an animal to the uropathogenic E. coli
XX strain J96.
XX
XX Sequence 314 BP; 105 A; 56 C; 60 G; 92 T; 1 other;
SQ

```

Query Match 5.9%; Score 30.6; DB 19; Length 314;
 Best Local Similarity 52.8%; Pred. No. 1.9;
 Matches 66; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```

QY 363 acatctgccattctctgttatgacagagcttggaacttaactgggactggggca 422
   ||||| || ||||| || || ||||| || || ||||| || || ||||| || ||
Db 84 aacatcagacgactctctcgttagaaaagctttattaggtttacaaacacaatgta 143
   ||||| || || || || || || || || || || || || || || || || || ||
QY 423 aagtcacatccacctttatcacatgaattgctgaagagcgcttttaaaactggagtgtg 482
   ||||| || || || || || || || || || || || || || || || || || ||
Db 144 cagtgattcaaacctctctattatagctcgttttagagactactggaagatggtgagtg 203
   ||||| || || || || || || || || || || || || || || || || || ||
QY 483 cattg 487
   |||||
Db 204 cattg 208

```

RESULT 12
 A15186
 ID A15186 standard; DNA; 11165 BP.
 AC A15186;
 XX
 DT 04-SEP-2000 (first entry)
 DE DNA encoding Escherichia coli virulence proteins.
 KW Virulence protein; tata; tatB; tatC; tatE; mdg; creC; recG; yggN;
 KW eck1; iroC; iroE; mtg2; msl; vaccine; infection;
 KW Gram negative bacterium; ss.
 XX
 OS Escherichia coli.
 XX
 FH Key Location/Qualifiers
 FT CDS 2..1099

```

FT /*tag= a
FT /product= "virulence protein"
FT /note= "encodes Y93237; no termination codon given"
FT 1102..1488
FT /*tag= b
FT /product= "virulence protein"
FT /note= "encodes Y93238"
FT 1573..1896
FT /*tag= c
FT /product= "virulence protein"
FT /note= "encodes Y93239"
FT 1939..2196
FT /*tag= d
FT /product= "virulence protein"
FT /note= "encodes Y93240; no termination codon given"
FT 2198..2533
FT /*tag= e
FT /product= "virulence protein"
FT /note= "encodes Y93241"
FT 2613..3041
FT /*tag= f
FT /product= "virulence protein"
FT /note= "encodes Y93242; no termination codon given"
FT 3054..3410
FT /*tag= g
FT /product= "virulence protein"
FT /note= "encodes Y93243"
FT 3460..3705
FT /*tag= h
FT /product= "virulence protein"
FT /note= "encodes Y93244"
FT 3791..4837
FT /*tag= i
FT /product= "virulence protein"
FT /note= "encodes Y93245"
FT 4878..7802
FT /*tag= j
FT /product= "virulence protein"
FT /note= "encodes Y93246"
FT 7816..9483
FT /*tag= k
FT /product= "virulence protein"
FT /note= "encodes Y93247"
FT 9836..10084
FT /*tag= l
FT /product= "virulence protein"
FT /note= "encodes Y93248"
FT 10134..10430
FT /*tag= m
FT /product= "virulence protein"
FT /note= "encodes Y93249"
FT 10459..10779
FT /*tag= n
FT /product= "virulence protein"
FT /note= "encodes Y93250"
FT
XX WO200028038-A2.
XX
XX 18-MAY-2000.
XX
XX 09-NOV-1999; 99WO-GB03721.
XX
XX 09-NOV-1998; 98GB-0024569.
XX 09-NOV-1998; 98GB-0024570.
XX 17-DEC-1998; 98GB-0027814.
XX 17-DEC-1998; 98GB-0027815.
XX 17-DEC-1998; 98GB-0027816.
XX 17-DEC-1998; 98GB-0027818.
XX 13-JAN-1999; 99GB-0000708.
XX 13-JAN-1999; 99GB-0000710.
XX 13-JAN-1999; 99GB-0000711.
XX 28-JAN-1999; 99GB-0001915.
XX

```


XX 19-MAR-1999; 99US-0125362.
PR 10-DEC-1999; 99US-0169980.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen GA, Ruben SM, Komatsoulis G;
PI
XX WPI; 2000-579483/54.
DR P-PSDB; B39209.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
PT
XX Claim 1; Page 357-358; 434pp; English.
XX
XX The polynucleotide sequences given in C74223-C74279 encode the human
CC secreted proteins represented in B39179-B39226. Sequences B39227-B39308
CC are alternative proteins encoded by the genes, and also protein sequences
CC with which they share homology. The proteins have activities based on the
CC tissues and cells in which they are expressed. Examples of activities
CC include: immunosuppressive; antiarthritic; antirheumatic;
CC antiproliferative; cytostatic; cardiac; vasotropic; cerebroprotective;
CC antitropic; neuroprotective; antibacterial; virucide; fungicide; and
CC ophthalmological. The human secreted proteins, polynucleotides,
CC antagonists and agonists of the invention may be useful in the treatment,
CC prevention, and/or diagnosis of various disease, disorders and conditions
CC such as autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
CC ocular disorders e.g. corneal infection. The polypeptides can also be
CC used to aid wound healing and epithelial cell proliferation, to
CC regenerate tissues, maintain organs before transplantation, in chemotaxis
CC and as a food additive or preservative e.g. to increase storage
CC capabilities. Sequences C74214-C74222 and B39178 are used during the
CC isolation and characterisation of the genes of the invention.
XX
SQ Sequence 3161 BP; 1025 A; 485 C; 603 G; 1048 T; 0 other;

Query Match 5.9%; Score 30.4; DB 21; Length 3161;
Best Local Similarity 50.7%; Pred. No. 7.4;
Matches 73; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY 372 ccattcttctgttatgacagagctgtgaaactttaactgggactggggcaagcaatc 431
D 1716 ccatttggaaatcatcaagaagatgttaactctgcttttagcattgatgcaagttataa 1775
QY 432 ccacctttatacaatgaattgtgaaagagccttttaaaacttgagtgatgcttttta 491
D 1776 taactgacattataaaaaatcatttgggaagccttttgaattatattgataagggttaa 1835
QY 492 tgaaggaggcttccattgatgcc 515
D 1836 tggagtcctctctcctatgcatac 1859

RESULT 15
242012/C
ID 242012 standard; cDNA; 692 BP.
XX
AC 242012;
XX
XX 31-JAN-2000 (first entry)
DE Human endometrium tumour cDNA derived EST 32.
XX
KW Endometrium; human; tumour; cancer; anticancer; cytostatic; EST:
KW treatment; uterine; gene therapy; expressed sequence tag; ss.
XX
OS Homo sapiens.

XX
PN
XX
PD
XX
XX
PF 17-APR-1998; 98DE-1017948.
XX
PR 17-APR-1998; 98DE-1017948.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX
DR WPI; 1999-591957/51.
DR P-PSDB; Y60034, Y60035, Y60036.
XX
XX New nucleic acid sequences expressed in uterine cancer tissues, and
PT derived polypeptides, for treatment of uterine and endometrial cancer
PT and identification of therapeutic agents -
XX
XX Claim 3; Page 190; 444pp; German.
XX
XX This invention describes novel human nucleic acid (cDNA) sequences (A),
CC that are highly expressed in uterine tumour tissue and which have
CC anticancer and cytostatic activity. (A) are used (i) for recombinant
CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)
CC are used (i) to identify agents suitable for treatment of uterine or
CC endometrial cancer; (ii) directly for treating these forms of cancer
CC (including expression from gene therapy vectors) and (iii) for generation
CC of specific antibodies. (A) are identified by assembling ESTs (expressed
CC sequence tags) from a particular tissue type before comparison of
CC expression patterns. This allows a significantly longer fragment of the
CC gene to be revealed, so should reduce the number of failures associated
CC with the fact that ESTs from different libraries may represent different
CC parts of the same unknown gene, distorting the estimated frequency of
CC occurrence in a particular tissue. 241981-242121 represent EST fragments
CC derived from a human endometrium tumour cDNA library which encode the
CC protein sequences represented in Y59941-Y60328.
XX
SQ Sequence 692 BP; 146 A; 189 C; 197 G; 160 T; 0 other;

Query Match 5.8%; Score 30; DB 20; Length 692;
Best Local Similarity 57.4%; Pred. No. 4.5;
Matches 54; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY 1 tttaaagtgcgtctttagaagccactgcctcagcttctgagctcagataccagaagg 60
D 645 TTAAAGTAGCTGACATGAGAAAGCCCTTGTAAATCCGGGGGCAAGTGACCAGAGGA 586
QY 61 agtctggtacacagcatgataaaagacaatggga 94
D 585 AGACTTCTTGAGCTCTCTCTCAAGCCAAAGGGA 552

Search completed: June 21, 2001, 18:06:24
Job time: 13792 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 18:20:53 ; Search time 150.46 Seconds
(without alignments)
634.116 Million cell updates/sec

Title: US-09-445-201-1_COPY_10094_10608

Perfect score: 515

Sequence: 1 tttaaatgtgtgtcttag.....agggtcttcttgatcc 515

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	29.6	5.7	152331	4	US-09-128-155-16
C 2	29.2	5.7	3688	6	Sequence 16, Appl
C 3	28.8	5.6	2088	4	Patent No. 5248670-4
C 4	28.4	5.5	2017	1	Sequence 64, Appl
C 5	28.4	5.5	2017	1	Sequence 27, Appl
C 6	28.4	5.5	2017	1	Sequence 27, Appl
C 7	28.4	5.5	2017	1	Sequence 27, Appl
C 8	28.2	5.5	410	4	Sequence 27, Appl
C 9	28.2	5.5	751	4	Sequence 88, Appl
C 10	28.2	5.5	3773	4	Sequence 7, Appl
C 11	28	5.4	1143	2	Sequence 1, Appl
C 12	28	5.4	2369	2	Sequence 19, Appl
C 13	28	5.4	2369	3	Sequence 4, Appl
C 14	28	5.4	12588	2	Sequence 4, Appl
C 15	27.8	5.4	376	4	Sequence 1, Appl
C 16	27.4	5.3	349	1	Sequence 3, Appl
C 17	27.4	5.3	2236	3	Sequence 3, Appl
C 18	27.4	5.3	2236	4	Sequence 11, Appl
C 19	27.4	5.3	4680	1	Sequence 11, Appl
C 20	27.4	5.3	4680	1	Sequence 1, Appl
C 21	27.4	5.3	4680	2	Sequence 1, Appl
C 22	27.4	5.3	4680	5	Sequence 1, Appl
C 23	27.4	5.3	4910	2	Sequence 1, Appl
C 24	27.4	5.3	4910	2	Sequence 2, Appl
C 25	27.2	5.3	710	1	Sequence 2, Appl
C 26	27.2	5.3	710	1	Sequence 22, Appl
C 27	27.2	5.3	710	1	Sequence 22, Appl

```

C 28 27.2 5.3 710 1 US-08-483-554B-22 Sequence 22, Appl
C 29 27.2 5.3 710 1 US-08-488-011B-22 Sequence 22, Appl
C 30 27.2 5.3 710 4 US-08-850-727-22 Sequence 22, Appl
C 31 27.2 5.3 710 5 PCT-US95-10202-22 Sequence 22, Appl
C 32 27.2 5.3 710 5 PCT-US95-10203-22 Sequence 22, Appl
C 33 27.2 5.3 710 5 PCT-US95-10220-22 Sequence 22, Appl
C 34 27.2 5.3 2610 2 US-08-989-386-2 Sequence 1, Appl
C 35 27.2 5.3 5656 1 US-08-425-061-1 Sequence 1, Appl
C 36 27.2 5.3 5656 1 US-08-825-886-1 Sequence 1, Appl
C 37 27.2 5.3 5689 1 US-08-425-061-3 Sequence 3, Appl
C 38 27.2 5.3 5689 1 US-08-825-886-3 Sequence 3, Appl
C 39 27.2 5.3 5707 1 US-08-425-061-11 Sequence 11, Appl
C 40 27.2 5.3 5707 1 US-08-825-886-11 Sequence 11, Appl
C 41 27.2 5.3 5709 1 US-08-425-061-2 Sequence 2, Appl
C 42 27.2 5.3 5709 1 US-08-425-061-7 Sequence 7, Appl
C 43 27.2 5.3 5709 1 US-08-425-061-8 Sequence 8, Appl
C 44 27.2 5.3 5709 1 US-08-425-061-9 Sequence 9, Appl
C 45 27.2 5.3 5709 1 US-08-825-886-2 Sequence 2, Appl

```

ALIGNMENTS

RESULT 1

```

US-09-128-155-16/c
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

```

```

Query Match 5.7%; Score 29.6; DB 4; Length 152331;
Best Local Similarity 51.5%; Pred.No. 41;
Matches 68; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```

```

QY 158 agggagtgttcattcaatcagcaatttagtcagatctgtgcatctctgtcttacaagaaa 217
      ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25835 ATGGTGAAGTCAATTTCTTGTGTCATGTGAAGGAGAAATGCTGTACTATTATTTTAAACAGA 25776
      ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 218 tgtcaatgggctgagatcatcagatggaggttcacatcggtttcaatgtccccgtatcctt 277
      ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25775 AAGGAGAGGGTTTATGGAGATTACAGACAGGATTTCTTGATAGCAATGATTGTCATCAAT 25716
      ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 278 ttgtaagacctt 289
      ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25715 TTAGAAGATTTT 25704
      ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 2

```

5248670-4/c
; Patent No. 5248670
; APPLICANT: DRAPER, KENNETH G.; ECKER, DAVID J.; MTRABELLI,
; CHRISTOPHER K.; CROOKE, STANLEY T.

```



```
RESULT 5
US-08-459-967-27/c
; Sequence 27, Application US/08459967
; Patent No. 5622841
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan P.
; APPLICANT: Outtrup, Helle
; APPLICANT: Dambmann, Claus
; APPLICANT: Aaslyng, Dorrit
; TITLE OF INVENTION: ALKALINE PROTEASE
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5622841o No. 5622841disk of No. 5622841th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,967
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/434,255
; FILING DATE: 03-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Agis Dr., Cheryl H.
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 3764.400-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2017 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-459-967-27

Query Match 5.5%; Score 28.4; DB 1; Length 2017;
Best Local Similarity 54.9%; Pred. No. 9.5;
Matches 56; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 355 gggttgaccatctgccattctctctgttatgacagagcttgaaactttaactggga 414
| ||||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 423 GCTTTGTTTCAGAGCAGAACCTTGTCCTTTTGGATACACCATCTTTGAATTGACGATGA 364
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 415 ctgggggcaagtcacatccacacctttatacaatgaattgctga 456
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 363 CTTACCGGATTTCGAAGCGCCTTTATCCAAAGTTTGTGA 322
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
US-08-460-327-27/c
; Sequence 27, Application US/08460327
; Patent No. 5622850
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan P.
; APPLICANT: Outtrup, Helle
; APPLICANT: Dambmann, Claus
; APPLICANT: Aaslyng, Dorrit
; TITLE OF INVENTION: ALKALINE PROTEASE
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5622850o No. 5622850disk of No. 5622850th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
```

```
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,327
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/434,255
; FILING DATE: 03-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Agis Dr., Cheryl H.
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 3764.400-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2017 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-460-327-27

Query Match 5.5%; Score 28.4; DB 1; Length 2017;
Best Local Similarity 54.9%; Pred. No. 9.5;
Matches 56; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 355 gggttgaccatctgccattctctctgttatgacagagcttgaaactttaactggga 414
| ||||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 423 GCTTTGTTTCAGAGCAGAACCTTGTCCTTTTGGATACACCATCTTTGAATTGACGATGA 364
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 415 ctgggggcaagtcacatccacacctttatacaatgaattgctga 456
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 363 CTTACCGGATTTCGAAGCGCCTTTATCCAAAGTTTGTGA 322
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
US-08-459-871-27/c
; Sequence 27, Application US/08459871
; Patent No. 5650326
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan P.
; APPLICANT: Outtrup, Helle
; APPLICANT: Dambmann, Claus
; APPLICANT: Aaslyng, Dorrit
; TITLE OF INVENTION: ALKALINE PROTEASE
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5650326o No. 5650326disk of No. 5650326th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,871
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: US 08/434,255
; FILING DATE: 03-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Agis Dr., Cheryl H.
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 3764,400-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2017 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; CLONE: 1651564
; US-08-459-871-27

Query Match          5.5%; Score 28.4; DB 1; Length 2017;
Best Local Similarity 54.9%; Pred. No. 9.5;
Matches 56; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 355 gtttggaccatcgcacattctctcttattgacagagcttggaactttaactggga 414
DB 423 GCTTGTTCAGACGACGACCTTGTGCGCTTTTGGATACACCATCTTTGAATTGACGATGA 364

QY 415 ctggggcaagtcacatccacaccttatacaatgaattgctga 456
DB 363 CTCACCGGATTCGAAAGCGCTTTATCCAAAGTTTGTGA 322

RESULT 8
US-09-276-531-88
; Sequence 88, Application US/09276531
; Patent No. 6183968
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Reddy, Roopa
; APPLICANT: Guegler, Karl J.
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING
; NUMBER OF SEQUENCES: 134
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/276,531
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/079,677
; FILING DATE: March 27, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lynn E. Murry, Ph.D.
; REGISTRATION NUMBER: 42,918
; REFERENCE/DOCKET NUMBER: PA-0008 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
```

```
; INFORMATION FOR SEQ ID NO: 88:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 410 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSTUT08
; CLONE: 1651564
; US-09-276-531-88

Query Match          5.5%; Score 28.2; DB 4; Length 410;
Best Local Similarity 53.1%; Pred. No. 4.5;
Matches 60; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 402 actttaactgggactggggcaaaagtcaatccacacctttatacaatgaattgctgaagagg 461
DB 71 AATCTGTGCTGCCAAGATAAAGTCAACTTCATATTTATATAGTTATATATTCGCGAGA 130

QY 462 ccttttaaaacttgagtgatgcatgtttatggaaggccttccctattggatc 514
DB 131 TCTGAGAAAGATGGACGTTACACTGATCTTCTATGTTCTTTTATTAGTTTC 183

RESULT 9
US-09-130-242-7
; Sequence 7, Application US/09130242B
; Patent No. 6194558
; GENERAL INFORMATION:
; APPLICANT: Gianturco, S.H.
; APPLICANT: Bradley, W.A.
; TITLE OF INVENTION: DNA Encoding Human Monocyte-Macrophage Aoplipoprotein
; FILE REFERENCE: D5880
; CURRENT APPLICATION NUMBER: US/09/130,242B
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: MS WORD, Macintosh OS 8.5
; SEQ ID NO 7
; LENGTH: 751
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-130-242-7

Query Match          5.5%; Score 28.2; DB 4; Length 751;
Best Local Similarity 54.3%; Pred. No. 6.4;
Matches 57; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 34 agcttctgcagctcagataccaaaggaagctctgttacacagcatgataaaagacaatggg 93
DB 285 agggcctggagatgacagaagacatgaagtgggagctcagctgtagaacagacctgggg 344

QY 94 acgggggtcacagtggctccctcccttttcagggggtatggagacga 138
DB 345 ctggggagatggcagctccctccctgggtcccaagcagagagggcagga 389

RESULT 10
US-09-130-242-1
; Sequence 1, Application US/09130242B
; Patent No. 6194558
; GENERAL INFORMATION:
; APPLICANT: Gianturco, S.H.
; APPLICANT: Bradley, W.A.
; TITLE OF INVENTION: DNA Encoding Human Monocyte-Macrophage Aoplipoprotein
; FILE REFERENCE: D5880
; CURRENT APPLICATION NUMBER: US/09/130,242B
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: MS WORD, Macintosh OS 8.5
```

APPLICANT: Bandman, Olga


```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,950
; FILING DATE: 05-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,279
; FILING DATE: 21-MARCH-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/033,181
; FILING DATE: 16-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/032,252
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0670K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)852-9196
; TELEFAX: (650)496-1204
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 376 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 78..374
; US-08-985-950-3

```

```

Query Match      5.4%; Score 27.8; DB 4; Length 376;
Best Local Similarity 54.4%; Pred. No. 5.9;
Matches 56; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 285 accctgaagtgcacagcaggaacacaggaactccaccctggtgccgtgaattgcagag 344
    ||||| ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 AACTTCAGGGGAAATAGAGAGAGAGGGGATGTCGATGGAGCCACCCTGGACTCCAGAG 182
    ||||| ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 345 ctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 387
    ||||| ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 CAGCGTTCAGGTGGTTGACCCCAAGCCCAATTTTCTGTGTGT 139
    ||||| ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: June 21, 2001, 18:21:24
Job time: 17477 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 15:33:59 ; Search time 4579.66 Seconds
(without alignments)
1063.009 Million cell updates/sec

Title: US-09-445-201-1-copy_10094_10608

Perfect score: 515

Sequence: 1 tttaaatgtgtgtcttag.....agggttcctattggtacc 515

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: gb_est1:*
- 2: gb_est2:*
- 3: gb_est3:*
- 4: gb_est4:*
- 5: gb_est5:*
- 6: gb_est6:*
- 7: gb_est7:*
- 8: gb_est8:*
- 9: gb_est9:*
- 10: gb_est10:*
- 11: gb_est11:*
- 12: gb_est12:*
- 13: gb_est13:*
- 14: gb_est14:*
- 15: gb_est15:*
- 16: gb_est16:*
- 17: gb_est17:*
- 18: gb_est18:*
- 19: gb_est19:*
- 20: gb_est20:*
- 21: gb_est21:*
- 22: gb_est22:*
- 23: gb_est23:*
- 24: gb_est24:*
- 25: gb_est25:*
- 26: gb_est26:*
- 27: gb_est27:*
- 28: gb_est28:*
- 29: gb_est29:*
- 30: gb_est30:*
- 31: gb_est31:*
- 32: gb_est32:*
- 33: gb_est33:*
- 34: gb_est34:*
- 35: gb_est35:*
- 36: gb_est36:*
- 37: gb_est37:*
- 38: gb_est38:*
- 39: gb_est39:*
- 40: gb_est40:*
- 41: gb_est41:*
- 42: gb_est42:*
- 43: gb_est43:*
- 44: gb_est44:*
- 45: gb_est45:*
- 46: gb_est46:*
- 47: gb_est47:*
- 48: gb_est48:*
- 49: gb_est49:*
- 50: gb_est50:*
- 51: gb_est51:*
- 52: gb_est52:*
- 53: gb_est53:*
- 54: gb_est54:*
- 55: gb_est55:*
- 56: gb_est56:*
- 57: gb_est57:*
- 58: gb_est58:*
- 59: gb_est59:*
- 60: gb_est60:*
- 61: gb_est61:*
- 62: gb_est62:*
- 63: gb_est63:*
- 64: gb_est64:*
- 65: gb_est65:*
- 66: gb_est66:*
- 67: gb_est67:*
- 68: gb_est68:*
- 69: gb_est69:*
- 70: gb_est70:*
- 71: gb_est71:*
- 72: gb_est72:*
- 73: gb_est73:*
- 74: gb_est74:*
- 75: gb_est75:*
- 76: gb_est76:*
- 77: gb_est77:*
- 78: gb_est78:*
- 79: gb_est79:*
- 80: gb_est80:*
- 81: gb_est81:*
- 82: gb_est82:*
- 83: gb_est83:*
- 84: gb_est84:*
- 85: gb_est85:*
- 86: gb_est86:*
- 87: gb_est87:*
- 88: gb_est88:*
- 89: gb_est89:*
- 90: gb_est90:*
- 91: gb_est91:*
- 92: gb_est92:*
- 93: gb_est93:*
- 94: gb_est94:*
- 95: gb_est95:*
- 96: gb_est96:*
- 97: gb_est97:*
- 98: gb_est98:*
- 99: gb_est99:*
- 100: gb_est100:*
- 101: gb_est101:*
- 102: gb_est102:*
- 103: gb_est103:*
- 104: gb_est104:*
- 105: gb_est105:*
- 106: gb_est106:*
- 107: gb_est107:*
- 108: gb_est108:*
- 109: gb_est109:*
- 110: gb_est110:*
- 111: gb_est111:*
- 112: gb_est112:*
- 113: gb_est113:*
- 114: gb_est114:*
- 115: gb_est115:*
- 116: gb_est116:*
- 117: gb_est117:*
- 118: gb_est118:*
- 119: gb_est119:*
- 120: gb_est120:*
- 121: gb_est121:*
- 122: gb_est122:*
- 123: gb_est123:*
- 124: gb_est124:*
- 125: gb_est125:*
- 126: gb_est126:*
- 127: gb_est127:*
- 128: gb_est128:*
- 129: gb_est129:*
- 130: gb_est130:*
- 131: gb_est131:*
- 132: gb_est132:*
- 133: gb_est133:*
- 134: gb_est134:*
- 135: gb_est135:*
- 136: gb_est136:*
- 137: gb_est137:*
- 138: gb_est138:*
- 139: gb_est139:*
- 140: gb_est140:*
- 141: gb_est141:*
- 142: gb_est142:*
- 143: gb_est143:*
- 144: gb_est144:*
- 145: gb_est145:*
- 146: gb_est146:*
- 147: gb_est147:*
- 148: gb_est148:*
- 149: gb_est149:*
- 150: gb_est150:*
- 151: gb_est151:*
- 152: gb_est152:*
- 153: gb_est153:*
- 154: gb_est154:*
- 155: gb_est155:*
- 156: gb_est156:*
- 157: gb_est157:*
- 158: gb_est158:*
- 159: gb_est159:*
- 160: gb_est160:*
- 161: gb_est161:*
- 162: gb_est162:*
- 163: gb_est163:*
- 164: gb_est164:*
- 165: gb_est165:*
- 166: gb_est166:*
- 167: gb_est167:*
- 168: gb_est168:*
- 169: gb_est169:*
- 170: gb_est170:*
- 171: gb_est171:*
- 172: gb_est172:*
- 173: gb_est173:*
- 174: gb_est174:*
- 175: gb_est175:*
- 176: gb_est176:*
- 177: gb_est177:*
- 178: gb_est178:*
- 179: gb_est179:*
- 180: gb_est180:*
- 181: gb_est181:*
- 182: gb_est182:*
- 183: gb_est183:*
- 184: gb_est184:*
- 185: gb_est185:*
- 186: gb_est186:*
- 187: gb_est187:*
- 188: gb_est188:*
- 189: gb_est189:*
- 190: gb_est190:*
- 191: gb_est191:*
- 192: gb_est192:*
- 193: gb_est193:*
- 194: gb_est194:*
- 195: gb_est195:*
- 196: gb_est196:*
- 197: gb_est197:*
- 198: gb_est198:*
- 199: gb_est199:*
- 200: gb_est200:*
- 201: gb_est201:*
- 202: gb_est202:*
- 203: gb_est203:*
- 204: gb_est204:*
- 205: gb_est205:*
- 206: gb_est206:*
- 207: gb_est207:*
- 208: gb_est208:*
- 209: gb_est209:*
- 210: gb_est210:*
- 211: gb_est211:*
- 212: gb_est212:*
- 213: gb_est213:*
- 214: gb_est214:*
- 215: gb_est215:*
- 216: gb_est216:*
- 217: gb_est217:*
- 218: gb_est218:*
- 219: gb_est219:*
- 220: gb_est220:*
- 221: gb_est221:*
- 222: gb_est222:*
- 223: gb_est223:*
- 224: gb_est224:*
- 225: gb_est225:*
- 226: gb_est226:*
- 227: gb_est227:*
- 228: gb_est228:*
- 229: gb_est229:*
- 230: gb_est230:*
- 231: gb_est231:*
- 232: gb_est232:*
- 233: gb_est233:*
- 234: gb_est234:*
- 235: gb_est235:*
- 236: gb_est236:*
- 237: gb_est237:*
- 238: gb_est238:*
- 239: gb_est239:*
- 240: gb_est240:*
- 241: gb_est241:*
- 242: gb_est242:*
- 243: gb_est243:*
- 244: gb_est244:*
- 245: gb_est245:*
- 246: gb_est246:*
- 247: gb_est247:*
- 248: gb_est248:*
- 249: gb_est249:*
- 250: gb_est250:*
- 251: gb_est251:*
- 252: gb_est252:*
- 253: gb_est253:*
- 254: gb_est254:*
- 255: gb_est255:*
- 256: gb_est256:*
- 257: gb_est257:*
- 258: gb_est258:*
- 259: gb_est259:*
- 260: gb_est260:*
- 261: gb_est261:*
- 262: gb_est262:*
- 263: gb_est263:*
- 264: gb_est264:*
- 265: gb_est265:*
- 266: gb_est266:*
- 267: gb_est267:*
- 268: gb_est268:*
- 269: gb_est269:*
- 270: gb_est270:*
- 271: gb_est271:*
- 272: gb_est272:*
- 273: gb_est273:*
- 274: gb_est274:*
- 275: gb_est275:*
- 276: gb_est276:*
- 277: gb_est277:*
- 278: gb_est278:*
- 279: gb_est279:*
- 280: gb_est280:*
- 281: gb_est281:*
- 282: gb_est282:*
- 283: gb_est283:*
- 284: gb_est284:*
- 285: gb_est285:*
- 286: gb_est286:*
- 287: gb_est287:*
- 288: gb_est288:*
- 289: gb_est289:*
- 290: gb_est290:*
- 291: gb_est291:*
- 292: gb_est292:*
- 293: gb_est293:*
- 294: gb_est294:*
- 295: gb_est295:*
- 296: gb_est296:*
- 297: gb_est297:*
- 298: gb_est298:*
- 299: gb_est299:*
- 300: gb_est300:*
- 301: gb_est301:*
- 302: gb_est302:*
- 303: gb_est303:*
- 304: gb_est304:*
- 305: gb_est305:*
- 306: gb_est306:*
- 307: gb_est307:*
- 308: gb_est308:*
- 309: gb_est309:*
- 310: gb_est310:*
- 311: gb_est311:*
- 312: gb_est312:*
- 313: gb_est313:*
- 314: gb_est314:*
- 315: gb_est315:*
- 316: gb_est316:*
- 317: gb_est317:*
- 318: gb_est318:*
- 319: gb_est319:*
- 320: gb_est320:*
- 321: gb_est321:*
- 322: gb_est322:*
- 323: gb_est323:*
- 324: gb_est324:*
- 325: gb_est325:*
- 326: gb_est326:*
- 327: gb_est327:*
- 328: gb_est328:*
- 329: gb_est329:*
- 330: gb_est330:*
- 331: gb_est331:*
- 332: gb_est332:*
- 333: gb_est333:*
- 334: gb_est334:*
- 335: gb_est335:*
- 336: gb_est336:*
- 337: gb_est337:*
- 338: gb_est338:*
- 339: gb_est339:*
- 340: gb_est340:*
- 341: gb_est341:*
- 342: gb_est342:*
- 343: gb_est343:*
- 344: gb_est344:*
- 345: gb_est345:*
- 346: gb_est346:*
- 347: gb_est347:*
- 348: gb_est348:*
- 349: gb_est349:*
- 350: gb_est350:*
- 351: gb_est351:*
- 352: gb_est352:*
- 353: gb_est353:*
- 354: gb_est354:*
- 355: gb_est355:*
- 356: gb_est356:*
- 357: gb_est357:*
- 358: gb_est358:*
- 359: gb_est359:*
- 360: gb_est360:*
- 361: gb_est361:*
- 362: gb_est362:*
- 363: gb_est363:*
- 364: gb_est364:*
- 365: gb_est365:*
- 366: gb_est366:*
- 367: gb_est367:*
- 368: gb_est368:*
- 369: gb_est369:*
- 370: gb_est370:*
- 371: gb_est371:*
- 372: gb_est372:*
- 373: gb_est373:*
- 374: gb_est374:*
- 375: gb_est375:*
- 376: gb_est376:*
- 377: gb_est377:*
- 378: gb_est378:*
- 379: gb_est379:*
- 380: gb_est380:*
- 381: gb_est381:*
- 382: gb_est382:*
- 383: gb_est383:*
- 384: gb_est384:*
- 385: gb_est385:*
- 386: gb_est386:*
- 387: gb_est387:*
- 388: gb_est388:*
- 389: gb_est389:*
- 390: gb_est390:*
- 391: gb_est391:*
- 392: gb_est392:*
- 393: gb_est393:*
- 394: gb_est394:*
- 395: gb_est395:*
- 396: gb_est396:*
- 397: gb_est397:*
- 398: gb_est398:*
- 399: gb_est399:*
- 400: gb_est400:*
- 401: gb_est401:*
- 402: gb_est402:*
- 403: gb_est403:*
- 404: gb_est404:*
- 405: gb_est405:*
- 406: gb_est406:*
- 407: gb_est407:*
- 408: gb_est408:*
- 409: gb_est409:*
- 410: gb_est410:*
- 411: gb_est411:*
- 412: gb_est412:*
- 413: gb_est413:*
- 414: gb_est414:*
- 415: gb_est415:*
- 416: gb_est416:*
- 417: gb_est417:*
- 418: gb_est418:*
- 419: gb_est419:*
- 420: gb_est420:*
- 421: gb_est421:*
- 422: gb_est422:*
- 423: gb_est423:*
- 424: gb_est424:*
- 425: gb_est425:*
- 426: gb_est426:*
- 427: gb_est427:*
- 428: gb_est428:*
- 429: gb_est429:*
- 430: gb_est430:*
- 431: gb_est431:*
- 432: gb_est432:*
- 433: gb_est433:*
- 434: gb_est434:*
- 435: gb_est435:*
- 436: gb_est436:*
- 437: gb_est437:*
- 438: gb_est438:*
- 439: gb_est439:*
- 440: gb_est440:*
- 441: gb_est441:*
- 442: gb_est442:*
- 443: gb_est443:*
- 444: gb_est444:*
- 445: gb_est445:*
- 446: gb_est446:*
- 447: gb_est447:*
- 448: gb_est448:*
- 449: gb_est449:*
- 450: gb_est450:*
- 451: gb_est451:*
- 452: gb_est452:*
- 453: gb_est453:*
- 454: gb_est454:*
- 455: gb_est455:*
- 456: gb_est456:*
- 457: gb_est457:*
- 458: gb_est458:*
- 459: gb_est459:*
- 460: gb_est460:*
- 461: gb_est461:*
- 462: gb_est462:*
- 463: gb_est463:*
- 464: gb_est464:*
- 465: gb_est465:*
- 466: gb_est466:*
- 467: gb_est467:*
- 468: gb_est468:*
- 469: gb_est469:*
- 470: gb_est470:*
- 471: gb_est471:*
- 472: gb_est472:*
- 473: gb_est473:*
- 474: gb_est474:*
- 475: gb_est475:*
- 476: gb_est476:*
- 477: gb_est477:*
- 478: gb_est478:*
- 479: gb_est479:*
- 480: gb_est480:*
- 481: gb_est481:*
- 482: gb_est482:*
- 483: gb_est483:*
- 484: gb_est484:*
- 485: gb_est485:*
- 486: gb_est486:*
- 487: gb_est487:*
- 488: gb_est488:*
- 489: gb_est489:*
- 490: gb_est490:*
- 491: gb_est491:*
- 492: gb_est492:*
- 493: gb_est493:*
- 494: gb_est494:*
- 495: gb_est495:*
- 496: gb_est496:*
- 497: gb_est497:*
- 498: gb_est498:*
- 499: gb_est499:*
- 500: gb_est500:*
- 501: gb_est501:*
- 502: gb_est502:*
- 503: gb_est503:*
- 504: gb_est504:*
- 505: gb_est505:*
- 506: gb_est506:*
- 507: gb_est507:*
- 508: gb_est508:*
- 509: gb_est509:*
- 510: gb_est510:*
- 511: gb_est511:*
- 512: gb_est512:*
- 513: gb_est513:*
- 514: gb_est514:*
- 515: gb_est515:*
- 516: gb_est516:*
- 517: gb_est517:*
- 518: gb_est518:*
- 519: gb_est519:*
- 520: gb_est520:*
- 521: gb_est521:*
- 522: gb_est522:*
- 523: gb_est523:*
- 524: gb_est524:*
- 525: gb_est525:*
- 526: gb_est526:*
- 527: gb_est527:*
- 528: gb_est528:*
- 529: gb_est529:*
- 530: gb_est530:*
- 531: gb_est531:*
- 532: gb_est532:*
- 533: gb_est533:*
- 534: gb_est534:*
- 535: gb_est535:*
- 536: gb_est536:*
- 537: gb_est537:*
- 538: gb_est538:*
- 539: gb_est539:*
- 540: gb_est540:*
- 541: gb_est541:*
- 542: gb_est542:*
- 543: gb_est543:*
- 544: gb_est544:*
- 545: gb_est545:*
- 546: gb_est546:*
- 547: gb_est547:*
- 548: gb_est548:*
- 549: gb_est549:*
- 550: gb_est550:*
- 551: gb_est551:*
- 552: gb_est552:*
- 553: gb_est553:*
- 554: gb_est554:*
- 555: gb_est555:*
- 556: gb_est556:*
- 557: gb_est557:*
- 558: gb_est558:*
- 559: gb_est559:*
- 560: gb_est560:*
- 561: gb_est561:*
- 562: gb_est562:*
- 563: gb_est563:*
- 564: gb_est564:*
- 565: gb_est565:*
- 566: gb_est566:*
- 567: gb_est567:*
- 568: gb_est568:*
- 569: gb_est569:*
- 570: gb_est570:*
- 571: gb_est571:*
- 572: gb_est572:*
- 573: gb_est573:*
- 574: gb_est574:*
- 575: gb_est575:*
- 576: gb_est576:*
- 577: gb_est577:*
- 578: gb_est578:*
- 579: gb_est579:*
- 580: gb_est580:*
- 581: gb_est581:*
- 582: gb_est582:*
- 583: gb_est583:*
- 584: gb_est584:*
- 585: gb_est585:*
- 586: gb_est586:*
- 587: gb_est587:*
- 588: gb_est588:*
- 589: gb_est589:*
- 590: gb_est590:*
- 591: gb_est591:*
- 592: gb_est592:*
- 593: gb_est593:*
- 594: gb_est594:*
- 595: gb_est595:*
- 596: gb_est596:*
- 597: gb_est597:*
- 598: gb_est598:*
- 599: gb_est599:*
- 600: gb_est600:*
- 601: gb_est601:*
- 602: gb_est602:*
- 603: gb_est603:*
- 604: gb_est604:*
- 605: gb_est605:*
- 606: gb_est606:*
- 607: gb_est607:*
- 608: gb_est608:*
- 609: gb_est609:*
- 610: gb_est610:*
- 611: gb_est611:*
- 612: gb_est612:*
- 613: gb_est613:*
- 614: gb_est614:*
- 615: gb_est615:*
- 616: gb_est616:*
- 617: gb_est617:*
- 618: gb_est618:*
- 619: gb_est619:*
- 620: gb_est620:*
- 621: gb_est621:*
- 622: gb_est622:*
- 623: gb_est623:*
- 624: gb_est624:*
- 625: gb_est625:*
- 626: gb_est626:*
- 627: gb_est627:*
- 628: gb_est628:*
- 629: gb_est629:*
- 630: gb_est630:*
- 631: gb_est631:*
- 632: gb_est632:*
- 633: gb_est633:*
- 634: gb_est634:*
- 635: gb_est635:*
- 636: gb_est636:*
- 637: gb_est637:*
- 638: gb_est638:*
- 639: gb_est639:*
- 640: gb_est640:*
- 641: gb_est641:*
- 642: gb_est642:*
- 643: gb_est643:*
- 644: gb_est644:*
- 645: gb_est645:*
- 646: gb_est646:*
- 647: gb_est647:*
- 648: gb_est648:*
- 649: gb_est649:*
- 650: gb_est650:*
- 651: gb_est651:*
- 652: gb_est652:*
- 653: gb_est653:*
- 654: gb_est654:*
- 655: gb_est655:*
- 656: gb_est656:*
- 657: gb_est657:*
- 658: gb_est658:*
- 659: gb_est659:*
- 660: gb_est660:*
- 661: gb_est661:*
- 662: gb_est662:*
- 663: gb_est663:*
- 664: gb_est664:*
- 665: gb_est665:*
- 666: gb_est666:*
- 667: gb_est667:*
- 668: gb_est668:*
- 669: gb_est669:*
- 670: gb_est670:*
- 671: gb_est671:*
- 672: gb_est672:*
- 673: gb_est673:*
- 674: gb_est674:*
- 675: gb_est675:*
- 676: gb_est676:*
- 677: gb_est677:*
- 678: gb_est678:*
- 679: gb_est679:*
- 680: gb_est680:*
- 681: gb_est681:*
- 682: gb_est682:*
- 683: gb_est683:*
- 684: gb_est684:*
- 685: gb_est685:*
- 686: gb_est686:*
- 687: gb_est687:*
- 688: gb_est688:*
- 689: gb_est689:*
- 690: gb_est690:*
- 691: gb_est691:*
- 692: gb_est692:*
- 693: gb_est693:*
- 694: gb_est694:*
- 695: gb_est695:*
- 696: gb_est696:*
- 697: gb_est697:*
- 698: gb_est698:*
- 699: gb_est699:*
- 700: gb_est700:*
- 701: gb_est701:*
- 702: gb_est702:*
- 703: gb_est703:*
- 704: gb_est704:*
- 705: gb_est705:*
- 706: gb_est706:*
- 707: gb_est707:*
- 708: gb_est708:*
- 709: gb_est709:*
- 710: gb_est710:*
- 711: gb_est711:*
- 712: gb_est712:*
- 713: gb_est713:*
- 714: gb_est714:*
- 715: gb_est715:*
- 716: gb_est716:*
- 717: gb_est717:*
- 718: gb_est718:*
- 719: gb_est719:*
- 720: gb_est720:*
- 721: gb_est721:*
- 722: gb_est722:*
- 723: gb_est723:*
- 724: gb_est724:*
- 725: gb_est725:*
- 726: gb_est726:*
- 727: gb_est727:*
- 728: gb_est728:*
- 729: gb_est729:*
- 730: gb_est730:*
- 731: gb_est731:*
- 732: gb_est732:*
- 733: gb_est733:*
- 734: gb_est734:*
- 735: gb_est735:*
- 736: gb_est736:*
- 737: gb_est737:*
- 738: gb_est738:*
- 739: gb_est739:*
- 740: gb_est740:*
- 741: gb_est741:*
- 742: gb_est742:*
- 743: gb_est743:*
- 744: gb_est744:*
- 745: gb_est745:*
- 746: gb_est746:*
- 747: gb_est747:*
- 748: gb_est748:*
- 749: gb_est749:*
- 750: gb_est750:*
- 751: gb_est751:*
- 752: gb_est752:*
- 753: gb_est753:*
- 754: gb_est754:*
- 755: gb_est755:*
- 756: gb_est756:*
- 757: gb_est757:*
- 758: gb_est758:*
- 759: gb_est759:*
- 760: gb_est760:*
- 761: gb_est761:*
- 762: gb_est762:*
- 763: gb_est763:*
- 764: gb_est764:*
- 765: gb_est765:*
- 766: gb_est766:*
- 767: gb_est767:*
- 768: gb_est768:*
- 769: gb_est769:*
- 770: gb_est770:*
- 771: gb_est771:*
- 772: gb_est772:*
- 773: gb_est773:*
- 774: gb_est774:*
- 775: gb_est775:*
- 776: gb

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi. 1 (bases 1 to 409)
REFERENCE AUTHORS	Newman, T., deBruijn, F. J., Green, P., Keegstra, K., Kende, H., McIntosh L., Ohlrooge, J., Raikhel, N., Somerville, S., Thomashow, M., Retzel E. and Somerville, C.
TITLE	Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones
JOURNAL MEDLINE	Plant Physiol. 106, 1241-1255 (1994)
COMMENT	95148729 Contact: Thomas Newman MSU-DOE Plant Research Laboratory Michigan State University MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E. Lansing, Mi Tel: 517-353-0854 Fax: 517-353-9168 Email: 22313tcn@ibm.cl.msu.edu Seq primer: T7 dye primer.


```

QY 341 agagctgtgtgtgtgtgtgtgaccatctgccatttcttctgttatgacagagctgtg 400
    || || || || || || || || || || || || || || || || || || || || ||
Db 415 AGCTCAGTGAGTGTGTTAATACCCATCTTCCAAAGCCCTCTGTCNATTGTTTCATC 356
    || || || || || || || || || || || || || || || || || || || || ||
QY 401 aactttaactgggactggggcaagtcacatccacacctttatacaatgaattgctgaag 460
    || || || || || || || || || || || || || || || || || || || || ||
Db 355 TCATCTCCATTCAGTCTCTGTCCTTGCCTGGAAAGTATATGTGATCATTTGGAGAGAA 296
    || || || || || || || || || || || || || || || || || || || || ||
QY 461 gccttttaaaactggagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 506
    || || || || || || || || || || || || || || || || || || || || ||
Db 295 GCATTCGCTTGTGGAATTTTCAGTGTGTTTTCATGTTGTTTTCCT 250
    || || || || || || || || || || || || || || || || || || || || ||

RESULT 9
LOCUS BG457226/c 648 bp mRNA EST 19-MAR-2001
DEFINITION NF101A05PL1F1035 Phosphate starved leaf Medicago truncatula cDNA
ACCESSION BG457226
VERSION BG457226.1 GI:13380551
KEYWORDS barrel medic.
SOURCE Medicago truncatula
ORGANISM Medicago truncatula
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
1 (bases 1 to 648)
Liu, J., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores
H.R., Inman, J.T., Weller, J.W., May, G.D., and Harrison, M.J.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula phosphate-starved leaf library
Unpublished (2000)
JOURNAL Contact: Harrison MJ
COMMENT The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7325
Fax: 580 221 7380
Email: mjharrison@noble.org
Insert Length: 648 Std Error: 0.00
Plate: 101 row: A column: 05
Seq primer: TCACACGAGAAACGCTATGAC.
FEATURES
source
1..648
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone_lib="Phosphate starved leaf"
/tissue_type="leaf"
/dev_stage="trifoliolate"
/note="Vector: Lambda Zap; At the trifoliolate stage, M.
truncatula plants were transplanted to phosphate-free sand
and grown for a further 30 days. During this 30 day
period, the plants were fertilized twice weekly with 1/2
Hoaglands solution containing only 20uM potassium
phosphate. RNA was prepared from above ground tissues."
BASE COUNT 191 a 162 c 133 g 161 t 1 others
ORIGIN

Query Match 6.6%; Score 33.8; DB 153; Length 648;
Best Local Similarity 57.8%; Pred. No. 17;
Matches 59; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 398 gtgaactttaactgggactggggcaagtcacatccacacctttatacaatgaattgctgaa 457
    || || || || || || || || || || || || || || || || || || || || ||
Db 613 GGGAGNTGAACGTGACGTCGGCGTAGTCAATGCGCATCTTCTCTCAAAATTTGACAGA 554
    || || || || || || || || || || || || || || || || || || || || ||
QY 458 gaggcccttttaaaactggagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 499
    || || || || || || || || || || || || || || || || || || || || ||
Db 553 GCCGCTGTATGAACCTCAAGGGTCTTCAATCTCGTCGAG 512
    || || || || || || || || || || || || || || || || || || || || ||

```

```

RESULT 10
LOCUS BF640846/c 654 bp mRNA EST 19-DEC-2000
DEFINITION NF058D11N1F1093 Insect herbivory Medicago truncatula cDNA clone
ACCESSION BF640846
VERSION BF640846.1 GI:11905004
KEYWORDS barrel medic.
SOURCE Medicago truncatula
ORGANISM Medicago truncatula
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
1 (bases 1 to 654)
Korth, K., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores
H.R., Inman, J.T., Weller, J.W. and May, G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula insect herbivory library
Unpublished (2000)
JOURNAL Contact: Korth K
COMMENT Dept. of Plant Pathology
University of Arkansas
217 Plant Science Building, Fayetteville, AR 72701, USA
Tel: 501 575 5191
Fax: 501 575 7601
Email: kkorth@comp.uark.edu
Insert Length: 654 Std Error: 0.00
Plate: 058 row: D column: 11
Seq primer: TCACACGAGAAACGCTATGAC.
FEATURES
source
1..654
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone_lib="Insect herbivory"
/tissue_type="local and systemic leaves"
/dev_stage="mature"
/note="Vector: Lambda Zap; Library was produced from fully
expanded M. truncatula leaves of plants fed upon by
Spodoptera exigua (beet armyworm) for 24 hours. Systemic
(undamaged leaves from injured plants) and wounded leaves
were harvested and pooled."
BASE COUNT 194 a 163 c 131 g 165 t 1 others
ORIGIN

Query Match 6.6%; Score 33.8; DB 151; Length 654;
Best Local Similarity 57.8%; Pred. No. 17;
Matches 59; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 398 gtgaactttaactgggactggggcaagtcacatccacacctttatacaatgaattgctgaa 457
    || || || || || || || || || || || || || || || || || || || || ||
Db 603 GGGAGNTGAACGTGACGTCGGCGTAGTCAATGCGCATCTTCTCTCAAAATTTGACAGA 544
    || || || || || || || || || || || || || || || || || || || || ||
QY 458 gaggcccttttaaaactggagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 499
    || || || || || || || || || || || || || || || || || || || || ||
Db 543 GCCGCTGTATGAACCTCAAGGGTCTTCAATCTCGTCGAG 502
    || || || || || || || || || || || || || || || || || || || || ||

RESULT 11
LOCUS BF642756/c 659 bp mRNA EST 19-DEC-2000
DEFINITION NF070F06IN1F1058 Insect herbivory Medicago truncatula cDNA clone
ACCESSION BF642756
VERSION BF642756.1 GI:11906914
KEYWORDS barrel medic.
SOURCE Medicago truncatula
ORGANISM Medicago truncatula
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
1 (bases 1 to 659)

```

REFERENCE	1 (bases 1 to 659)
AUTHORS	Korth, K., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores, H.R., Imman, J.T., Weller, J.W. and May, G.D.
TITLE	Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula insect herbivory library
JOURNAL	Unpublished (2000)
COMMENT	Contact: Korth K

```

Plate: 106   row: A   column: 08
Seq primer: TCACACAGGAACAGCTATGAC.
Location/Qualifiers
1..661
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF106A08PL"
FEATURES
source

```

	Query Match	6.6%	Score 33.8;	DB 151;	Length 659;
	Best Local Similarity	57.8%;	Pred. No. 17;		
	Matches 59;	Conservative 0;	Mismatches 43;	Indels 0;	Gaps 0;
QY 398	gtgaacttttaactggggactggggcgaagtgcaatccccaccctttatacaaatgaattgctgaa	457			
Db 589	GGGAAGNTGAACGTGTACTCGGGCGTAGTCAATGCCATCTTCTCCTCAAATTTGACAGA	530			
QY 458	gaggccctttaaacttggagtgtagattgttatggaaaggg	499			
Db 529	GCCGCTGATGAACCTCAAGGGTCCTTCAATCTCGTCGAG	488			

RESULT	13
LOCUS	BG457203/c
DEFINITION	BG457203 666 bp mRNA EST 19-MAR-2001 NF095F08PLIF1073 Phosphate starved leaf Medicago truncatula cDNA clone NF095F08PL 5', mRNA sequence.
ACCESSION	BG457203
VERSION	BG457203.1 GI:13380528
KEYWORDS	EST.
SOURCE	barrel medic.
ORGANISM	Medicago truncatula Eukaryota; Viridiplantae; Streptophyta; Embryophyta Tracheophyta;

RESULT	12
LOCUS	BG457415/c
DEFINITION	Phosphate starved leaf Medicago truncatula cDNA clone NF106A08L10P5, mRNA sequence.
ACCESSION	U09682
VERSION	1.0
KEYWORDS	EST.
SOURCE	Medicago truncatula barrel medic.
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

REFERENCE	1 (pages 1 to 666)
AUTHORS	Liu, J., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores H.R., Inman, J.T., Weller, J.W., May, G.D. and Harrison, M.J.
TITLE	Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula phosphate-starved leaf library
JOURNAL	Unpublished (2000)
COMMENT	Contact: Harrison MJ Plant Biology Division The Samuel Roberts Noble Foundation 2510 Sam Noble Parkway, Ardmore, OK 73402, USA Tel: 580 221 7325

Fri Jun 22 10:02:14 2001

us-09-445-201-1-copy_6036_6959.rge

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 18:37:34 ; Search time 8015.43 Seconds
(without alignments)
1783.086 Million cell updates/sec

Title: US-09-445-201-1-COPY_6036_6959
Perfect score: 924
Sequence: 1 gaagttcaaacaccgaatgt.....tgccagcgaggtgcagg 924

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba1:
2: gb_ba2:
3: gb_ba3:
4: gb_in1:
5: gb_in2:
6: gb_in3:
7: gb_om:
8: gb_ov:
9: gb_pat1:
10: gb_pat2:
11: gb_ph:
12: gb_pl1:
13: gb_pl2:
14: gb_pl3:
15: gb_pl4:
16: em_ba1:
17: em_ba2:
18: em_fun:
19: em_htgo_hum:
20: em_htgo_inv:
21: em_htgo_rod:
22: em_htg_hum1:
23: em_htg_hum2:
24: em_htg_hum3:
25: em_htg_hum4:
26: em_htg_hum5:
27: em_htg_hum6:
28: em_htg_hum7:
29: em_htg_hum8:
30: em_htg_inv1:
31: em_htg_inv2:
32: em_htg_other:
33: em_htg_rod:
34: em_hum1:
35: em_hum2:
36: em_hum3:
37: em_hum4:
38: em_hum5:
39: em_hum6:
40: em_hum7:
41: em_in:
42: em_om:
43: em_or:

|||||tgggc 855
GAGAACTGGGC 915

44: em_ov:
45: em_pat:
46: em_ph:
47: em_pl:
48: em_ro:
49: em_sy:
50: em_sy:
51: em_un:
52: em_vi:
53: gb_sts1:
54: gb_sts2:
55: gb_sts3:
56: gb_sy:
57: gb_un:
58: gb_vil:
59: gb_vil:
60: gb_vil:
61: gb_vil:
62: gb_vil:
63: gb_vil:
64: gb_vil:
65: gb_vil:
66: gb_vil:
67: gb_vil:
68: gb_vil:
69: gb_vil:
70: gb_vil:
71: gb_vil:
72: gb_vil:
73: gb_vil:
74: gb_vil:
75: gb_vil:
76: gb_vil:
77: gb_vil:
78: gb_vil:
79: gb_vil:
80: gb_vil:
81: gb_vil:
82: gb_vil:
83: gb_vil:
84: gb_vil:
85: gb_vil:
86: gb_vil:
87: gb_vil:
88: gb_vil:
89: gb_vil:
90: gb_vil:
91: gb_vil:
92: gb_vil:
93: gb_vil:
94: gb_vil:
95: gb_vil:
96: gb_vil:
97: gb_vil:
98: em_ba3:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	924	100.0	939	94	AF153057 Mus muscu
2	924	100.0	12845	9	AX002124 Sequence
3	441.2	47.7	550	94	X89777 M.musculus
4	220.6	23.9	4790	95	S53103 NYK/FLK-1-
5	169	18.3	5470	9	AR068047 Sequence
6	169	18.3	5470	94	X70842 M.musculus
7	164.4	17.8	4016	95	U93307 Rattus norv
8	164.4	17.8	5892	95	U93306 Rattus norv

9	161	17.4	5391	94	MMFLK1	X59397 Mouse Flk-1
10	154.6	16.7	5406	9	AR005213	Sequence
11	154.6	16.7	5406	9	AR071705	Sequence
12	154.6	16.7	5406	10	I25171	Sequence 5
13	154.6	16.7	5406	10	I40602	Sequence 5
14	110	11.9	1267	93	HSDNAKDR	X89776 H.sapiens D
15	72	7.8	214348	66	AC021220	Homo sapi
16	71.4	7.7	5830	88	AF035121	Homo sapi
17	70.6	7.6	2264	9	AR030848	Sequence
18	70.6	7.6	2264	10	I82807	Sequence 16
19	50	5.4	3009	9	AX060542	Sequence
20	47	5.1	41622	3	SCD25	Alil18514 Streptomy
21	46.4	5.0	393	7	AF083095	Felis cat
22	44	4.8	16074	68	AC023682	AC023682 Drosophil
23	43.8	4.7	11958	1	AE004596	Pseudomon
24	42.4	4.6	1837	1	AF010151	Pseudomon
25	42.4	4.6	76201	85	AC005182	Homo sapi
26	42	4.5	168798	63	AC012683	Homo sapi
27	42	4.5	196579	74	AC069546	Homo sapi
28	41.6	4.5	178907	77	AC087726	Chlamydom
29	40.8	4.4	130278	86	AC005664	Homo sapi
30	40.6	4.4	191073	82	AP000752	Homo sapi
31	40.6	4.4	204842	82	AP001855	Homo sapi
32	40.2	4.4	155092	62	AC012073	Homo sapi
33	40	4.3	3897	8	CHKH33A1	MI1392 Chicken his
34	39.8	4.3	181715	78	AC090442	Homo sapi
35	39.4	4.3	5992	85	AB046862	Homo sapi
36	39.4	4.3	6636	91	AY004227	Homo sapi
37	39.4	4.3	7766	91	AY004226	Homo sapi
38	39.4	4.3	8089	89	AF311855	Homo sapi
39	39.4	4.3	8788	88	AF082075	Homo sapi
40	39.4	4.3	142300	66	AC020929	Homo sapi
41	39.4	4.3	163024	79	AL157939	Homo sapi
42	39.4	4.3	164652	66	AC021625	Homo sapi
43	39.4	4.3	166684	62	AC012386	Homo sapi
44	39.4	4.3	207375	81	AL392043	Homo sapi
45	39.2	4.2	876646	92	HS1048E9	Z997114 Human DNA s

ALIGNMENTS

RESULT 1	AF153057	939 bp	DNA	ROD	22-MAY-2000
LOCUS	Mus musculus tyrosine kinase FLK-1 gene, promoter and mRNA, partial sequence.				
DEFINITION	AF153057				
ACCESSION	AF153057.1 GI:7963622				
VERSION	house mouse.				
KEYWORDS	Mus musculus				
SOURCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
ORGANISM	1 (bases 1 to 939)				
REFERENCE	Kappel,A., Risau,W. and Breier,G.				
AUTHORS	Prerequisite role of SCL/Tal-1, GATA and Ets transcription factor				
TITLE	binding sites for the in vivo function of Flk-1 gene regulatory elements				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 939)				
AUTHORS	Kappel,A., Risau,W. and Breier,G.				
TITLE	Direct Submission				
JOURNAL	Submitted (21-MAY-1999) Molecular Cell Biology, Max-Planck, Institut fuer Physiologische und Klinische Forschung, Parkstrasse 1, Bad Nauheim 61231, Germany				
FEATURES	Location/Qualifiers				
source	1..939				
	/organism="Mus musculus"				
	/strain="129Sv"				
	/db_xref="taxon:10090"				
	/tissue_type="liver"				
	/dev_stage="fetus"				

promoter	1..640	Query Match	100.0%	Score 924;	DB 94;	Length 939;
mRNA	641..>939	Best Local Similarity	100.0%;	Pred. No. 2e-232;		
BASE COUNT	188 a 271 c 286 g 194 t	Matches 924;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
ORIGIN						
QY	1	gaagtcaaacacgaaatgtcttcttagggtaatacgttaacttcggacatttaagatt	60			
DB	16	GAAGTTCAACACGGAATGCTTCTTAGGGTAATACAGTAACCTCGGACGATTAAAGTT	75			
QY	61	gccagatgacagagaaacagtagagcgttgcaacctggataagcgcctacttctaa	120			
DB	76	GCAGATGACGAGAAACAGTAGAGCGGTGGCAACCTGGATAAGCGCTATCTCTAA	135			
QY	121	ttaaacatttcagagcggcggtggtggtggtggtggtggtggtggtggtggtggt	180			
DB	136	TTAAACATTCAGACGGGCGGGGATGCGGTGGCCAAAGCACCATATAAAACAACTTC	195			
QY	181	asgtactgacaaactcaactgcaagtttggcccgaggtacatcttaggttcagggttctt	240			
DB	196	AAGTACTGACCACTCACATGCAAGTTTGTGCCCGGATACATCTAGGTTTTCAGGGTTCTT	255			
QY	241	gtcttcactgacaaactcggcggtggtggtggtggtggtggtggtggtggtggtggt	300			
DB	256	GTCCTCATCTCTCCAACTCGGCGGGGATTTTGTGCTTGGGACTTTTCAGTTCAGAGGGC	315			
QY	301	gaagagagttctgcaacttcagagctcctaatgagggcgagtggtggtggtggtggtggt	360			
DB	316	GAAGAGAGTTCTGCACCTTCGAGGCTCTTAATGAGGCGCAGTGGGCTCGTGTCTCTGTT	375			
QY	361	gatgttccagagttctggtggggcagcaagtgtctcagagccattactggtacatttt	420			
DB	376	GATGCTTCCAGGTTCTGGGGGACGACAGTGTCTCAGAGCCCATTAATGGCTACATTTT	435			
QY	421	acttccacagaaacagcgtcgtccagatttgcctcagatgcgacttgcgcgcgcgcgcgc	480			
DB	436	ACTTCCACAGAAACGAGCTGCTCCAGATTGCTCTCAGATCGACTTGCCTCCCGCCGC	495			
QY	481	acagttccgggttagtggtgggtggtggtggtggtggtggtggtggtggtggtggtggt	540			
DB	496	ACAGTTCCGGGTAGTGGGGAGTGGGGAGTGGGGAGTGGGGAGTGGGGAGTGGGGAGT	555			
QY	541	agtggggcggtggcgagcag	600			
DB	556	AGTGGGGCGGTGGCGGACGAGGAGTCCCTCCCGCTCCCGCTCCCGCTCCCGCTCCCGCT	615			
QY	601	attcgttagtgtagcggcggtcctcttcttgcctgagtcctcagagaccccaagagagag	660			
DB	616	ATTCCGTAGTGTACCGCGGCTCTCTTCTGCCCTGAGTCTCTCAGAGACCCCAAGAGAG	675			
QY	661	taagctgttcttcttagatcgc	720			
DB	676	TAAGCTGTGTTCTTCTTAGATCGCGGACGCGCTACCGCGGAGGACTGAAAGCCAGACTG	735			
QY	721	gtcccgagc	780			
DB	736	TGTCGCGACCGCGGATTAACCTGGCTGACCCGATTCGCGGACACCGCTCGAGCGCGGC	795			
QY	781	tggagcagggcggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt	840			
DB	796	TGSAGCAGGGCGCGGTGCCCGGCTCTCCCGGCTCTCCCGGCTCTCCCGGCTCTCCCGGCT				
QY	841	gctctgtgactcttcttggggcagggcagggcagggcagggcagggcagggcagggcagggc				
DB	856	GCTCTGTGACTTCTTTGGCGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG				
QY	901	tctgtgcccagc				

Db 916 TCTGTGCCCCAGCGGAGGTGCAGG 939

RESULT 2
AX002124
LOCUS AX002124 12845 bp DNA PAT 10-MAR-2000
DEFINITION Sequence 1 from Patent WO9855638.
ACCESSION AX002124
VERSION AX002124.1 GI:7241839
KEYWORDS
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 12845)
AUTHORS Breier,G. and Risau,W.
TITLE REGULATORY SEQUENCES CAPABLE OF CONFERRING EXPRESSION OF A HETEROLOGOUS DNA SEQUENCE IN ENDOTHELIAL CELLS IN VIVO AND USES THEREOF
JOURNAL Patent: WO 9855638-A 1 10-DEC-1998;
MAX PLANCK GESELLSCHAFT (DE); BREIER GEORG (DE)
FEATURES
Location/Qualifiers
1..12845
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 3321 a 2893 c 2863 g 3723 t 45 others
ORIGIN

Query Match 100.0%; Score 924; DB 9; Length 12845;
Best Local Similarity 100.0%; Pred. No. 1.3e-232; Indels 0; Gaps 0;
Matches 924; Conservative 0; Mismatches 0;

QY 1 gaagttcacacgaatgtctttagggcctaatacaggttaacttcgacgatttaagtt 60
|||||
Db 6036 GAAGTTCACACCGAATGCTCTTAGGGCTAATCAGGTAACCTCGGACGATTAAAGTT 6095
|||||
QY 61 gccatgagacagaaaaacagtagagcgttggcaacctggataagcgcctatctctaa 120
|||||
Db 6096 GCCAGATGACGAGAAAACAGTAGAGGCGTTGGCAACCTGGGATAGCGCTATCTTCTAA 6155
|||||
QY 121 ttaaacattcagaggggggggagtcggtggcgaagcaccataaaacaaacttcc 180
|||||
Db 6156 TTAACAACTTACAGCGGGCGGGGATGCGGTGGCCAAAGCACCATAAAACAACTTCC 6215
|||||
QY 181 aagttactgacaaactcactcgaagttgtgccccagtagtacatctaggttcagggttctt 240
|||||
Db 6216 AAGTACTGACCAACTCACTGCAAGTTTGTGCCCGAGTACATCTAGTTTCAGGGGTCTT 6275
|||||
QY 241 gtcttcagtcctcccaactgcggcgagattttgttccttgggaactttcagtcgagcggc 300
|||||
Db 6276 GTCTTCATGCTCCCAACTCGCGGGCGGATTTTGTGGTCCCTTGGGACTTTCAGTGCAGCGGC 6335
|||||
QY 301 gaagagagttctgacttcaggtcctcctaataagggcgagtcggcctcgtgtttctggt 360
|||||
Db 6336 GAAGAGAGTTCTGCACTTGCAGGCTCCCTAATCAGGCGCAGTGGGCCCTGCTGTTCTGTT 6395
|||||
QY 361 gatgttccccaggttgcctg999ggcgagagtgctctcagagccccattactggctacatttt 420
|||||
Db 6396 GATGTTTCCCAAGTTGCTGTGGGGGAGCAAGTGTCTCAGAGCCCATTAAGTGTACATTTT 6455
|||||
QY 421 acttcacacagaaacccagctcgtccagatttctctcagatgcgacttgcgcgcggc 480
|||||
Db 6456 ACTTCCACAGAAACCGAGCTCGTCCAGATTGTCTCAGATGGGACTTTCGCGCCCGGC 6515
|||||
QY 481 acagttccggggtagtg999ggagtg99cg9tggaaacgggaaacccaaacctggtatcc 540
|||||
Db 6516 ACAGTTCGGGGTAGTGGGGGAGTGGGGTGGGAAACCGGAAACCCAAACCTGGTATCC 6575
|||||
QY 541 aatggggggcggtg99cgagcagggaggtccccacccctcccggttaatgaccccgcccc 600
|||||
Db 6576 AGTGGGGGGCGTGGCGGACGAGGAGTCCCCACCCCTCCCGGTAAATGACCCCGCCGCC 6635
|||||
QY 601 attcgtagttgtgtagcggcgctctcttcttctgctgagtcagtcagggagagagag 660
|||||

Db 6636 ATTGCGTAGTGTAGCGCGGCTCTCTTCTGCCCTAGTCTCAGACCCCAAGAGAG 6695
|||||
QY 661 taagctgtgttcttagtagtcg99ggaccgctaccg99caggactgaaagccagactg 720
|||||
Db 6696 TAAAGCTGTGTTTCCTTAGATCGCGGACCGCTACCCGGCAGGACTGAAAGCCAGACTG 6755
|||||
QY 721 tgtcccgagcgggataacctgctgacccgattccg99gagacacccgctgagcgcggc 780
|||||
Db 6756 TGTCCCGCAGCGGGATAACCTGGCTGACCGGATTCGCGGACACCCGCTGCAGCCGGC 6815
|||||
QY 781 tggagccagggcgccggtgcccgcgctctcccgcgtcttgcgctg99ggg9gcatacc 840
|||||
Db 6816 TGGAGCCAGGCGCGGCTGCCCGCGCTCTCCCGGCTTTCGCGTGGGGGCGCATACC 6875
|||||
QY 841 gcctctgtgactcttctg99ggcagggcagggagagagctgtgctgagaaactgggc 900
|||||
Db 6876 GCCTCTGTGACTTCTTTGCGGCGCAGGAGGAGGAGTCTGTGCTGAGAACTGGC 6935
|||||
QY 901 tctgtgcccagcgcgaggtgcagg 924
|||||
Db 6936 TCTGTGCCAGCGGAGGTGCAGG 6959
|||||

RESULT 3
MMDNAKDR
LOCUS MMDNAKDR 550 bp DNA ROD 10-DEC-1995
DEFINITION M.musculus DNA for 5'UTR of receptor tyrosine.
ACCESSION X89777
VERSION X89777.1 GI:1019394
KEYWORDS KDR/flk-1 gene.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 550)
AUTHORS Patterson,C., Perrella,M.A., Hsieh,C.M., Yoshizumi,M., Lee,M.E. and Haber,E.
TITLE Cloning and functional analysis of the promoter for KDR/flk-1, a receptor for vascular endothelial growth factor
J. Biol. Chem. 270 (39), 23111-23118 (1995)
MEDLINE 96032749
REFERENCE 2 (bases 1 to 550)
AUTHORS Patterson,C.
TITLE Direct Submission
JOURNAL Submitted (19-JUL-1995) C. Patterson, Harvard School of Public Health, Cardiovascular Biology Lab, 677 Huntington Avenue, Building 2 Room 113a, Boston MA 02115, USA
FEATURES
Location/Qualifiers
1..550
/organism="Mus musculus"
/db_xref="taxon:10090"
promoter 1..295
/gene="KDR/flk-1"
gene 1..550
/gene="KDR/flk-1"
CDS 504..>550
/gene="KDR/flk-1"
/codon_start=1
/product="vascular endothelial growth factor (VEGF) receptor"
/protein_id="CAA61917.1"
/db_xref="GI:1129146"
/db_xref="SWISS-PROT:P35918"
/translation="NESKALLAVALWFCV"
BASE COUNT 91 a 173 c 185 g 101 t
ORIGIN

Query Match 47.7%; Score 441.2; DB 94; Length 550;
Best Local Similarity 98.4%; Pred. No. 2.1e-105;
Matches 498; Conservative 0; Mismatches 3; Indels 5; Gaps 5;

RESULT 9

MMFLK1 5391 bp mRNA ROD 06-NOV-1991
 Mouse Flk-1 mRNA for a tyrosine kinase receptor.
 X59397
 VERSION X59397.1 GI:50976
 KEYWORDS Flk-1 gene; tyrosine kinase receptor.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 5391)
 Jordan,C.T.
 Direct Submission
 Submitted (09-MAY-1991) C.T. Jordan, Princeton University, Dept.
 Molecular Biology, Washington Rd, Princeton, New Jersey 08544, USA
 2 (bases 1 to 5391)
 Matthews,W., Jordan,C.T., Gavin,M., Jenkins,N.A., Copeland,N.G. and
 Lemischka,I.R.
 TITLE A receptor tyrosine kinase cDNA isolated from a population of
 enriched primitive hematopoietic cells and exhibiting close genetic
 linkage to c-kit
 Proc. Natl. Acad. Sci. U.S.A. 88 (20), 9026-9030 (1991)
 92020984
 FEATURES Location/Qualifiers
 1..5391
 /organism="Mus musculus"
 /strain="C3H/He"
 /db_xref="taxon:10090"
 /chromosome="5"
 1..5391
 /gene="Flk-1"
 /evidence=experimental
 1..5391
 /gene="Flk-1"
 208..431
 /gene="Flk-1"
 /codon_start=1
 /product="tyrosine kinase receptor"
 /protein_id="CAA42040.1"
 /db_xref="GI:50977"
 /db_xref="MGD:MG1:95555"
 /db_xref="SWISS-PROT:P35918"
 /translation="MESKALLAVLWFCVETRAASVGLPGDFLHPKLSQKDLITIL
 ANTIQTCRQDLDLWPNARDESEVLTECGGDSIFPKTLTIPRVVGNDTGA
 YKSYRDYDIASVYVYRDYSPFIASVSDHGIYITENKNTWVLPGRGISNLN
 VSLCARYPEKRFVDPGNRISWDSEIGFTLPSTYMSIYAGMVFCEAKINDETYSINLV
 VVGVYRIDVILSPDHEIELSAGEKLVNLTARTELNVLGDFETHSPSPKSHKKIVN
 RDYKFPFTVAKMFLSTLITISVKSDDGETCVASSGRMKNRTFFVHVKPFIAF
 GSGMKSLEATVGSQVRIPVLYLSYPADPDIKWYNGRPISNTYIMVDELTIMETVE
 RDAGNYTILNPTISMEKSHMVSILVNVPPQIGEKALISPDMSYQGTMTLTCVY
 ANPLHLHIQWYMLEACSYRPGQTSYACKERHVEFDQGNKLEVTKNQYALIEGK
 NKTSTVLIOQANYSALYKCAINKAGRGVISPVIRGPETVQPAAPQTEQESVS
 LLCTADRNTFENLWYKLSOATSVHMGESLTPVCKNLDAWLKNGTFESNTDILI
 VAFONASLQDGDYVCSAQDKTKRKHCLVQLIILERMAMPINLGNENQITIGETI
 EVTCPSAGNPFTIIFWQAEKRLNLEVLVGTVAIEMFFLLVILVTVTRANGRELKGT
 LGCAEAETFLIEGAQERLNVLLVGTVAIEMFFLLVILVTVTRANGRELKGT
 YLSIVMDPELPLDCERLPLDASKFEPRDLKPLGRGAFQVIEADAFIDK
 TATCTKAVKMLKEGATSEHRALMSKILIHIGHNLVNLGACFKPGPLAVIV
 EFSKFNLSLTYLKRNEFPVYKSGAREGKDYVGLSVDLKRLDSITSSQSSAS
 SGFVEEKSLSDVEEEASEELYKDFLTLEHLICYSFQVAKMEFLASRCLHRDLAAR
 NILLSEKNVVICDFGLARDYKDPDYVRKGDARPLKWPAPETIFDRVYITQSDWS
 FGLLWEIFSLGASYPGVKIDEEFCRLKEGTRMRAPDYTPPEYQTMIDCWHDEN
 QRPSFSELVHLNLLQAQDKDYIVLPMSETLSMEEDSGLSLPTSPVSCMBEE
 VCDPKFYDNTAGTSHYLNQSKRSRPVSKTFEDIPLPEPVKVIPODSDTDSGMVL
 ASEELKTLSDRNKLSRPGMMPKSRVSASEGNSNOTSGYQSGVHSDDTDTTVYSSD
 EAGLLKMDAAVHADSGTTLQITSLNGSGVPVPPPTPGNHERGA"
 208..264
 /gene="Flk-1"
 265..4308
 /gene="Flk-1"
 /product="tyrosine kinase receptor"
 1397 a 1299 c 1422 g 1273 t

source

mRNA

gene

CDS

sig_peptide

mat_peptide

BASE COUNT

ORIGIN

Query Match 17.4%; Score 161; DB 94; Length 5391;
 Best Local Similarity 98.1%; Pred. No. 7.2e-32;
 Matches 205; Conservative 0; Mismatches 0; Indels 4; Gaps 4;
 QY 718 ctgtgtcccgagccgggataaactgtgctgacccgattccgcgacacaccgtgcagccgc 777
 Db 1 CTGTGTCCCGCAGCC -GGATAACCTGGCTGACCCGATTCGCCGACACCCGTGCAGCCGC 59
 QY 778 ggttgagccagggcgccggtgccccgcgtctctccccggtcttgcgctggggggcgcat 837
 Db 60 GGCTGGAGCCAGGCGCGGCTGCCCGCTCTCCCCGGTCTTCCCGTGGGGGSC-CAT 118
 QY 838 accgcctctgtgactcttcttgcgggcccagggacggaggaagagctgtgctctgag-aact 896
 Db 119 ACCGCCCTCTGTGACTTCTTTTCGGGGCCAGGACGGAGAGAGTCTGTGCTGAGAACT 178
 QY 897 gggctctgtgcccac-gcgcgaggtgcagg 924
 Db 179 GGCTCTGTGCCAGGCGCGAGGTGCAGG 207

RESULT 10

AR005213
 LOCUS AR005213 5406 bp DNA PAT 04-DEC-1998
 DEFINITION Sequence 5 from patent US 5747651.
 ACCESSION AR005213
 VERSION AR005213.1 GI:3966092
 KEYWORDS Unknown.
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 5406)
 AUTHORS Lemischka,I.R.
 TITLE Antibodies against tyrosine kinase receptor flk-1
 JOURNAL Patent: US 5747651-A 5 05-MAY-1998;
 FEATURES Location/Qualifiers
 source 1..5406
 /organism="unknown"
 BASE COUNT 1412 a 1298 c 1423 g 1273 t
 ORIGIN

Query Match 16.7%; Score 154.6; DB 9; Length 5406;
 Best Local Similarity 96.2%; Pred. No. 3.5e-30;
 Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

QY 718 ctgtgtcccgagccgggataaactgtgctgacccgattccgcgacacaccgtgcagccgc 777
 Db 1 CTGTGTCCCGCAGCC -GGATAACCTGGCTGACCCGATTCGCCGACACCCGTGCAGCCGC 59
 QY 778 ggttgagccagggcgccggtgccccgcgtctctccccggtcttgcgctggggggcgcat 837
 Db 60 GGCTGGAGCCAGGCGCGGCTGCCCGCTCTCCCCGGTCTTCCGCTGGGGGSCCAT 118
 QY 838 accgcctctgtgactcttcttgcgggcccagggacggaggaagagctgtgctctgag-aact 896
 Db 119 ACCGCCCTCTGTGACTTCTTTTCGGGGCCAGGACGGAGAGAGTCTGTGCTGAGAACT 178
 QY 897 gggctctgtgcccac-gcgcgaggtgcagg 924
 Db 179 GGCTCTGTGCCAGGCGCGAGGTGCAGG 207

RESULT 11

AR071705
 LOCUS AR071705 5406 bp DNA PAT 18-FEB-2000
 DEFINITION Sequence 5 from patent US 5912133.
 ACCESSION AR071705
 VERSION AR071705.1 GI:7222593
 KEYWORDS

SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5406)
AUTHORS Lemischka, I.R.
TITLE Method for isolating stem cells expressing flk-1 receptors
JOURNAL Patent: US 5912133-A 5 15-JUN-1999;
FEATURES Location/Qualifiers
1..5406
/organism="unknown"
BASE COUNT 1412 a 1298 c 1423 g 1273 t
ORIGIN

Query Match 16.7%; Score 154.6; DB 9; Length 5406;
Best Local Similarity 96.2%; Pred. No. 3.5e-30;
Matches 20; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

QY 718 ctgtgtcccgagccgggataaactggtgacccgattccgcggagacacccgtgcagccgc 777
|||||
Db 1 CTGTGTCCCGCAGCC-GGATAAAGCTGGCTGACCGGATTCGCGGACACACCGCTGCAGCCGC 59
|||||

QY 778 ggctggagccagggcgccggtgcccgcgctctccccggtcttcccggtctgcgtgcggggcgcat 837
|||||
Db 60 GGCTGGAGCCAGGGCGCGGTG-CCCGCGCTCTCCCGGTCTTGGCGTGGCGGGGCGCAT 118
|||||

QY 838 accgcctctgtgactcttcttggcgccagggagcaggaagagtgctgtgctgag-aact 896
|||||
Db 119 ACCGCCCTCTGTGACTCTTTTGGCGGCCAGGGAGCGAGAGAGTCTGTGCTGTGAGAACT 178
|||||

QY 897 ggctctgtgcccc-a-gcgcgaggtgcagg 924
|||||
Db 179 GGGCTCTGTGTCGCCAGGCGCGAGGTGCAGG 207
|||||

RESULT 12
125171
LOCUS 125171 5406 bp DNA PAT 07-OCT-1996
DEFINITION Sequence 5 from patent US 5348065.
ACCESSION 125171
VERSION 125171.1 GI:1605041
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5406)
AUTHORS Lemischka, I.R.
TITLE Tyrosine kinase receptor human flk-2-specific antibodies
JOURNAL Patent: US 5548065-A 5 20-AUG-1996;
FEATURES Location/Qualifiers
1..5406
/organism="unknown"
BASE COUNT 1412 a 1298 c 1423 g 1273 t
ORIGIN

Query Match 16.7%; Score 154.6; DB 10; Length 5406;
Best Local Similarity 96.2%; Pred. No. 3.5e-30;
Matches 20; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

QY 718 ctgtgtcccgagccgggataaactggtgacccgattccgcggagacacccgtgcagccgc 777
|||||
Db 1 CTGTGTCCCGCAGCC-GGATAAAGCTGGCTGACCGGATTCGCGGACACACCGCTGCAGCCGC 59
|||||

QY 778 ggctggagccagggcgccggtgcccgcgctctccccggtcttcccggtctgcgtgcggggcgcat 837
|||||
Db 60 GGCTGGAGCCAGGGCGCGGTG-CCCGCGCTCTCCCGGTCTTGGCGTGGCGGGGCGCAT 118
|||||

QY 838 accgcctctgtgactcttcttggcgccagggagcaggaagagtgctgtgctgag-aact 896
|||||
Db 119 ACCGCCCTCTGTGACTCTTTTGGCGGCCAGGGAGCGAGAGAGTCTGTGCTGTGAGAACT 178
|||||

QY 897 ggctctgtgcccc-a-gcgcgaggtgcagg 924
|||||

Db 179 GGGCTCTGTGTCGCCAGGCGCGAGGTGCAGG 207
|||||

RESULT 13
140602
LOCUS 140602 5406 bp DNA PAT 13-MAY-1997
DEFINITION Sequence 5 from patent US 5621090.
ACCESSION 140602
VERSION 140602.1 GI:2082894
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5406)
AUTHORS Lemischka, I.R.
TITLE Nucleic acids encoding soluble human FLK-2 extracellular domain
JOURNAL Patent: US 5621090-A 5 15-APR-1997;
FEATURES Location/Qualifiers
1..5406
/organism="unknown"
BASE COUNT 1412 a 1298 c 1423 g 1273 t
ORIGIN

Query Match 16.7%; Score 154.6; DB 10; Length 5406;
Best Local Similarity 96.2%; Pred. No. 3.5e-30;
Matches 20; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

QY 718 ctgtgtcccgagccgggataaactggtgacccgattccgcggagacacccgtgcagccgc 777
|||||
Db 1 CTGTGTCCCGCAGCC-GGATAAAGCTGGCTGACCGGATTCGCGGACACACCGCTGCAGCCGC 59
|||||

QY 778 ggctggagccagggcgccggtgcccgcgctctccccggtcttcccggtctgcgtgcggggcgcat 837
|||||
Db 60 GGCTGGAGCCAGGGCGCGGTG-CCCGCGCTCTCCCGGTCTTGGCGTGGCGGGGCGCAT 118
|||||

QY 838 accgcctctgtgactcttcttggcgccagggagcaggaagagtgctgtgctgag-aact 896
|||||
Db 119 ACCGCCCTCTGTGACTCTTTTGGCGGCCAGGGAGCGAGAGAGTCTGTGCTGTGAGAACT 178
|||||

QY 897 ggctctgtgcccc-a-gcgcgaggtgcagg 924
|||||
Db 179 GGGCTCTGTGTCGCCAGGCGCGAGGTGCAGG 207
|||||

RESULT 14
HSDNAKDR
LOCUS HSDNAKDR 1267 bp DNA PRI 14-JAN-1996
DEFINITION H.sapiens DNA for 5'UTR of receptor tyrosine.
ACCESSION X89776
VERSION X89776.1 GI:1155008
KEYWORDS KDR/flk-1 gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1267)
AUTHORS Patterson, C., Perrella, M.A., Hsieh, C.M., Yoshizumi, M., Lee, M.E. and Haber, E.
TITLE Cloning and functional analysis of the promoter for KDR/flk-1, a receptor for vascular endothelial growth factor
JOURNAL J. Biol. Chem. 270 (39), 23111-23118 (1995)
MEDLINE 96032749
REFERENCE 2 (bases 1 to 1267)
AUTHORS Patterson, C.
TITLE Direct Submission
JOURNAL Submitted (19-JUL-1995) C. Patterson, Harvard School of Public Health, Cardiovascular Biology Lab, 677 Huntington Avenue, Building 2 Room 113a, Boston MA 02115, USA
REMARK Revised by [2]
REFERENCE 3 (bases 1 to 1267)
AUTHORS Patterson, C.

QY	760	ggacaccgctgcagccg-cgctggagcgaaggcgccgtgcccccgcgtctctccocggtc	818
Db	912	AGACGCCCTTCGAGCGCGGTTCGGCGCCGGTCCCTAGCCCTGTGCCTCAACTGTC	971
QY	819	ttagcctgcggg----gcgcataaccgcctctgtgaacttctttgcggcgccggagac--gg	872
Db	972	CATCGCTCGGGGGTGCCGCGAGTTCCACTCTCGCGCCTCTTCTTAGACAGGGCGCTGGG	1031
QY	873	agaaggagtctgtgcctagaactggctctgtgcccaagcgcgagggtgcagg	924
Db	1032	AGAAAGAACC GGCTCCC GAGTTC TGGCATTTCCGCGCGCTTCGAGGTGCAGG	1083

RESULT	15
AC021220/c	
LOCUS	AC021220 214348 bp DNA HTG
DEFINITION	Homo sapiens chromosome 4 clone RP11-530I17, WORKING DRAFT SEQUENCE, 38 unordered pieces.
ACCESSION	AC021220
VERSION	AC021220.4 GI:11178158
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.

REFERENCE
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 214348)
Waterston, R. H.
AUTHORS

```

JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 214348)
AUTHORS      Waterston,R.H.
TITLE        Direct Submission
JOURNAL      Submitted (15-JAN-2000) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
COMMENT      On Nov 15, 2000 this sequence version replaced gl:9309528.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC

```

```
----- Project Information -----
Center project name: H_NH0530I17
----- Summary Statistics -----
```

Sequencing vector: plasmid; 13%
Chemistry: Dye-primer ET; 87% of reads

Consensus quality: 195065 bases at least Q40
Consensus quality: 200600 bases at least Q30

Insert size: 101000; agarose-fp
Insert size: 210648; sum-of-contigs
Quality coverage: 4.96 in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 38 contigs. The true order of the pieces

- * arbitrary. Gaps between the contigs are represented as
- * runs of N, but the exact sizes of the gaps are unknown.
- * This record will be updated with the finished sequence

* be preserved.
* 1 1584: contig of 1584 bp in length
* 1585 1684: rap of unknown length

Accession	Gene	Accession	Gene
4195	4294: gap of unknown length	4195	4294: gap of unknown length
4295	6424: contig of 2130 bp in length	4295	6424: contig of 2130 bp in length
5425	5524: contig of 2130 bp in length	5425	5524: contig of 2130 bp in length

Abstract

.....

* 9845 9944: gap of unknown length
 * 9945 14101: contig of 4157 bp in length
 * 14102 14201: gap of unknown length
 * 14202 21279: contig of 7078 bp in length
 * 21280 21379: gap of unknown length
 * 21380 31346: contig of 9967 bp in length
 * 31347 31446: gap of unknown length
 * 31447 40526: contig of 9080 bp in length
 * 40527 40626: gap of unknown length
 * 40627 52862: contig of 12236 bp in length
 * 52863 52963: gap of unknown length
 * 52964 66090: contig of 13128 bp in length
 * 66091 66190: gap of unknown length
 * 66191 82659: contig of 16469 bp in length
 * 82660 82759: gap of unknown length
 * 82760 104797: contig of 22038 bp in length
 * 104798 104897: gap of unknown length
 * 104898 129319: contig of 24422 bp in length
 * 129320 129420: gap of unknown length
 * 129421 176511: contig of 47092 bp in length
 * 176512 176612: gap of unknown length
 * 176613 177666: contig of 1055 bp in length
 * 177667 177766: gap of unknown length
 * 177767 178937: contig of 1231 bp in length
 * 178938 179037: gap of unknown length
 * 179038 180242: contig of 1145 bp in length
 * 180243 180342: gap of unknown length
 * 180343 181442: contig of 1100 bp in length
 * 181443 181542: gap of unknown length
 * 181543 182642: contig of 1100 bp in length
 * 182643 182742: gap of unknown length
 * 182743 184165: contig of 1423 bp in length
 * 184166 184265: gap of unknown length
 * 184266 185112: contig of 1247 bp in length
 * 185113 185612: gap of unknown length
 * 185613 187121: contig of 1509 bp in length
 * 187122 187221: gap of unknown length
 * 187222 188524: contig of 1303 bp in length
 * 188525 188624: gap of unknown length
 * 188625 189782: contig of 1158 bp in length
 * 189783 189882: gap of unknown length
 * 189883 191312: contig of 1430 bp in length
 * 191313 191412: gap of unknown length
 * 191413 192837: contig of 1425 bp in length
 * 192838 192937: gap of unknown length
 * 192938 194117: contig of 1180 bp in length
 * 194118 194217: gap of unknown length
 * 194218 195948: contig of 1731 bp in length
 * 195949 196048: gap of unknown length
 * 196049 197744: contig of 1696 bp in length
 * 197745 197844: gap of unknown length
 * 197845 199489: contig of 1645 bp in length
 * 199490 199589: gap of unknown length
 * 199590 201304: contig of 1715 bp in length
 * 201305 201404: gap of unknown length
 * 201405 203316: contig of 1912 bp in length
 * 203317 203416: gap of unknown length
 * 203417 204490: contig of 1074 bp in length
 * 204491 204590: gap of unknown length
 * 204591 206676: contig of 2086 bp in length
 * 206677 206776: gap of unknown length
 * 206777 208803: contig of 2027 bp in length
 * 208804 208903: gap of unknown length
 * 208904 210597: contig of 1694 bp in length
 * 210598 210697: gap of unknown length
 * 210698 212097: contig of 1400 bp in length
 * 212098 212197: gap of unknown length
 * 212198 214346: contig of 2151 bp in length.

FEATURES
 source

1. .214348
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"

/clone="RP11-530117"
 1. .1584
 /note="assembly_name:Contig100"
 1685. .4194
 /note="assembly_name:Contig101"
 4295. .6424
 /note="assembly_name:Contig102"
 6525. .9844
 /note="assembly_name:Contig103"
 9945. .14101
 /note="assembly_name:Contig104"
 14202. .21279
 /note="assembly_name:Contig105"
 21380. .31346
 /note="assembly_name:Contig106"
 31447. .40526
 /note="assembly_name:Contig107"
 clone_end:77
 vector_side:left
 40627. .52862
 /note="assembly_name:Contig108"
 52963. .66090
 /note="assembly_name:Contig109"
 66191. .82659
 /note="assembly_name:Contig110"
 82760. .104797
 /note="assembly_name:Contig111"
 104898. .129319
 /note="assembly_name:Contig112"
 clone_end:SP6
 vector_side:right
 129420. .176511
 /note="assembly_name:Contig113"
 176612. .177666
 /note="assembly_name:Contig133"
 177667. .178997
 /note="assembly_name:Contig137"
 179098. .180242
 /note="assembly_name:Contig156"
 180343. .181442
 /note="assembly_name:Contig159"
 181543. .182642
 /note="assembly_name:Contig166"
 182743. .184165
 /note="assembly_name:Contig169"
 184266. .185512
 /note="assembly_name:Contig175"
 185613. .187121
 /note="assembly_name:Contig177"
 187222. .188524
 /note="assembly_name:Contig178"
 188625. .189782
 /note="assembly_name:Contig180"
 189883. .191312
 /note="assembly_name:Contig185"
 191413. .192837
 /note="assembly_name:Contig186"
 192938. .194117
 /note="assembly_name:Contig187"
 194218. .195948
 /note="assembly_name:Contig188"
 196049. .197744
 /note="assembly_name:Contig189"
 197845. .199489
 /note="assembly_name:Contig190"
 199590. .201304
 /note="assembly_name:Contig191"
 201405. .203316
 /note="assembly_name:Contig193"
 203417. .204490

Query Match 7.8%; Score 72; DB 66; Length 214348;
 Best Local Similarity 67.3%; Pred. No. 1.1e-08;

Search completed: June 21, 2001, 18:38:55
Job time: 18714 sec

Result No.	Query %			ID	Description
	Score	Match	Length		
1	924	100.0	12845	20	V84274
2	374.2	40.5	500	18	T60766
3	169	18.3	5470	15	Q64049
4	154.6	16.7	5404	17	X38735
5	154.6	16.7	5404	20	T377516
6	154.6	16.7	5406	13	Q29957
7	154.6	16.7	5406	14	Q35504
8	154.6	16.7	5406	14	O53251
9	154.6	16.7	5406	14	Q40916
10	154.6	16.7	5406	16	Q81014
11	154.6	16.7	5406	16	Q79070

FT /*tag= d

LE

Db	6756	tgctccgcgcgcgggataaacctggtgtgaccccgattccgcggacacccgctgcagccgcggc	6815
Qy	781	tggagccaggcgccgcgtgccccgcgtctccccgcgtcttcgcgtgcggggcgccataacc	840
Db	6816	tggagccaggcgccgcgtgccccgcgtctccccgcgtcttcgcgtgcggggcgccataacc	6875
Qy	841	gcctctgtgaactcttttcggggccaggagacggagagaagaagctctgtcctgagaactgggc	900
Db	6876	gcctctgtgaactcttttcggggccaggagacggagagaagaagctctgtcctgagaactgggc	6935
Qy	901	tctgtgcccagcgcgaggtgcagg	924
Db	6936	tctgtgcccagcgcgaggtgcagg	6959
RESULT 2			
T60766	ID	T60766 standard; DNA; 500 BP.	
T60766;	XX		
XX	AC		
XX	AC		
DT	15-OCT-1997	(first entry)	
DE	Mouse KDR/flk-1 promoter bases (-295) to 205.		
XX			
KW	KDR/flk-1; promoter; regulation; endothelial cell; transcription;		
KW	gene therapy; treatment; specific; ds.		
OS	Mus spp.		
XX			
Key	Location/Qualifiers		
misc_feature	39..44		
FT	/*tag= a		
FT	/note= "potential cis-acting element (E Box) "		
FT	52..57		
FT	/*tag= b		
FT	/note= "potential cis-acting element (Sp1) "		
FT	101..109		
FT	/*tag= d		
FT	/note= "potential cis-acting element (Ker) "		
FT	128..135		
FT	/*tag= e		
FT	/note= "potential cis-acting element (Ap2) "		
FT	143..157		
FT	/*tag= f		
FT	/note= "potential cis-acting element (NFKB/Ap2) "		
FT	172..177		
FT	/*tag= g		
FT	/note= "potential cis-acting element (Sp1) "		
FT	313..316		
FT	/*tag= h		
FT	/note= "potential cis-acting element (GATA) "		
XX	W09700957 -A1.		
PN			
PF	09-JAN-1997.		
PD			
XX			
XX	21-JUN-1996;	96WO-US10725.	
XX			
PPR	18-DEC-1995;	95US-0573692.	
PPR	23-JUN-1995;	95US-0494282.	
XX			
PA	(HARD) HARVARD COLLEGE.		
XX			
XX	Haber E, Lee M, Patterson-Winston C;		
PI			
XX	WPI; 1997-087384/08.		
DR			
XX			
PPT	KDR/flk-1 promoter sequence - useful for regulation of endothelial		
PPT	cell specific transcription of operably linked polypeptide-encoding		
PPT	sequences		
XX			
PS	Disclosure; Page 46; 70pp; English.		

DE Sequence of murine genomic foetal liver kinase 1 (Flk-1) gene.

XX Foetal liver kinase; receptor tyrosine kinase; ss.

XX Mus musculus.

XX Key Location/Qualifiers

FT CDS 286..4386

FT /*tag= a

XX W09411499-A.

XX 26-MAY-1994.

XX 15-NOV-1993; 93WO-EP03191.

XX 13-NOV-1992; 92US-0975750.

XX 26-MAR-1993; 93US-0038596.

XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX Millauer B, Risau W, Ullrich A;

XX WPI; 1994-183501/22.

XX P-PSDB; R54046.

XX DNA encoding Flk-1, a tyrosine kinase receptor for vascular

XX endothelial growth factor - used to express recombinant Flk-1 for

XX screening for ligands useful for modulating vasculogenesis and

XX angiogenesis e.g. for treating cancer

XX Disclosure; Page 50-56; 99pp; English.

XX A receptor tyrosine kinase cDNA, designated foetal liver kinase 1

XX (Flk-1), was cloned from mouse cell populations enriched for

XX haematopoietic stem and progenitor cells. The nucleotide coding

XX sequence and deduced AA sequence of the murine Flk-1 gene is

XX depicted in Q64049/R54046) and has been described in Matthews et al.,

XX 1991, Proc. Natl. Acad. Sci. USA, 88:9026-9030. The murine Flk-1

XX gene was isolated by performing a PCR using two degenerate oligo

XX primer pools that were designed on the basis of highly conserved

XX sequences within the kinase domain of receptor tyrosine kinases

XX (Hanks et al., 1988). Based on AA homology, this receptor is a member

XX of the type III subclass of RTKs (Ullrich and Schlessinger) which

XX contd. immunoglobulin-like repeats in their extracellular domains.

XX Sequence 5470 BP; 1417 A; 1324 C; 1440 G; 1289 T; 0 other;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

AC

XX

DT

XX

XX

DE

XX

KW

XX

KW

XX

KW

XX

KW

XX

OS

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

T38735;

11-DEC-1996 (first entry)

Murine foetal liver kinase 1 cDNA.

Murine; foetal liver kinase 1; flk-1; protein tyrosine kinase; monoclonal; antibody; extracellular domain; receptor assay; haematopoietic stem cell; ligand; stimulation; proliferation; differentiation; treatment; anaemia; bone marrow damage; cancer chemotherapy; radiation; ds.

Mus musculus.

Key Location/Qualifiers

CDS 208..4311

FT /*tag= a

FT sig_peptide

FT /*tag= b

FT mat_peptide

FT /*tag= c

XX US548065-A.

XX 20-AUG-1996.

XX 02-APR-1991; 91US-0679666.

XX 19-NOV-1992; 92US-0977451.

XX 02-APR-1991; 91US-0679666.

XX 28-JUN-1991; 91US-0728913.

XX 15-NOV-1991; 91US-0793065.

XX 24-DEC-1991; 91US-0813593.

XX 26-JUN-1992; 92US-0906397.

XX 12-NOV-1992; 92US-0975049.

XX 30-APR-1993; 93US-0055269.

XX 31-OCT-1994; 94US-0252517.

XX (UVP-) UNIV PRINCETON.

XX Lemischka IR;

XX WPI; 1996-392678/39.

XX P-PSDB; T38735.

XX Anti-foetal liver kinase 2 (flk-2) antibodies - useful in assays,

XX for isolating haematopoietic stem cells expressing receptor and for

XX obtaining ligands

XX Disclosure; Columns 51-62; 50pp; English.

XX The present sequence encodes murine foetal liver kinase 1 (flk-1),

XX a protein tyrosine kinase. Isolated antibodies, pref. monoclonal,

XX raised against the extracellular portion of flk-1 can be used to

XX assay for flk receptors on the surface of haematopoietic stem

XX cells, and to isolate positive cells. The antibodies can also

XX be used as, or to obtain ligands, which stimulate the proliferation

XX and/or differentiation of stem cells. The ligands can be used, e.g.

XX for treating anaemia, or bone marrow damage resulting from cancer

XX chemotherapy, or radiation.

XX Sequence 5404 BP; 1411 A; 1297 C; 1423 G; 1273 T; 0 other;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

DE

XX

KW

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

DE

XX

XX

XX

XX

XX

XX

XX

CC haematopoietic cells. Ligand binding to the TK may be prep'd. which
 CC can stimulate proliferation and/or differentiation of primitive
 CC haematopoietic cells in vivo. The ligands can stimulate the
 CC proliferation of additional primitive stem cells, differentiation into
 CC more mature progenitor cells, or both.
 CC See also Q29954-6.
 XX
 SQ Sequence 5406 BP; 1412 A; 1294 C; 1423 G; 1273 T; 0 other;

Query Match 16.7%; Score 154.6; DB 13; Length 5406;
 Best Local Similarity 96.2%; Pred. No. 4.7e-34;
 Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;
 QY 718 ctgtgtcccgagccgggataacctgtgtgacccgattcccgagacacccgtgcagccgc 777
 Db 1 ctgtgtcccgagcc-ggataacctgtgtgacccgattcccgagacacccgtgcagccgc 59
 QY 778 ggtgtgagccagggcgccggtgccccgctctccccggtcttctgctgagggcgcat 837
 Db 60 ggtgtgagccagggcgccggtg-ccccgctctccccggtcttctgctgagggcgcat 118
 QY 838 accgctctgtgactttcttgcggcgccagggagcagagagagtgctgtgctgag-aact 896
 Db 119 accgctctgtgactttcttgcggcgccagggagcagagagagtgctgtgctgagaaact 178
 QY 897 gggctctgtgcccc-gcgcgaggtgcagg 924
 Db 179 gggctctgtgccccgagcgaggtgcagg 207

RESULT 7

Q53504

ID Q53504 standard; cDNA; 5406 BP.

XX AC Q53504;

XX 27-JUN-1994 (first entry)

XX Murine flk-1 cDNA.

DE Receptor protein tyrosine kinase; pTK family; foetal liver kinase;

KW mflk; primitive; totipotent; haematopoietic cell; stem cell;

KW proliferation; stromal cell; ds.

XX Mus musculus.

FH Key Location/Qualifiers

FT CDS 208..4311

FT /*tag= a

FT /product= mflk-1

FT 208..264

FT /*tag= b

FT /note= "hydrophobic leader"

FT 265..4308

FT /*tag= c

XX US5270458-A.

XX 14-DEC-1993.

XX 02-APR-1991; 91US-0679666.

XX 02-APR-1991; 91US-0679666.

XX 28-JUN-1991; 91US-0728913.

XX 15-NOV-1991; 91US-0793065.

XX 24-DEC-1991; 91US-0813593.

XX 26-JUN-1992; 92US-0906397.

XX 12-NOV-1992; 92US-0975049.

XX 19-NOV-1992; 92US-0977451.

XX (UYPR-) UNIV PRINCETON.

XX PA

XX XX

PI Lemischka IR;

XX WPI; 1993-405021/50.

DR P-PSDB; R44996.

XX

PT Isolated nucleic acid molecules of hematopoietic stem cell

PT receptor flk-2 - encoding mammalian receptor protein tyrosine

PT kinases expressed in primitive haematopoietic cells

XX

PS Disclosure; Fig 2; 60pp; English.

XX

CC Nucleic acid sequences coding for murine flk-2 and specified

CC subfragments of it are claimed. The murine flk-1 coding sequence

CC (i.e. Q53504) is also disclosed. The flk polypeptides are

CC receptor protein tyrosine kinases which are expressed only in

CC primitive haematopoietic cells.

XX

SQ Sequence 5406 BP; 1412 A; 1298 C; 1423 G; 1273 T; 0 other;

Query Match

Best Local Similarity 16.7%; Score 154.6; DB 14; Length 5406;

Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

QY 718 ctgtgtcccgagccgggataacctgtgtgacccgattcccgagacacccgtgcagccgc 777

Db 1 ctgtgtcccgagcc-ggataacctgtgtgacccgattcccgagacacccgtgcagccgc 59

QY 778 ggtgtgagccagggcgccggtgccccgctctccccggtcttctgctgagggcgcat 837

Db 60 ggtgtgagccagggcgccggtg-ccccgctctccccggtcttctgctgagggcgcat 118

QY 838 accgctctgtgactttcttgcggcgccagggagcagagagagtgctgtgctgag-aact 896

Db 119 accgctctgtgactttcttgcggcgccagggagcagagagagtgctgtgctgagaaact 178

QY 897 gggctctgtgcccc-gcgcgaggtgcagg 924

Db 179 gggctctgtgccccgagcgaggtgcagg 207

RESULT 8

Q35251

ID Q35251 standard; cDNA; 5406 BP.

XX AC Q35251;

XX 25-JUN-1993 (first entry)

XX Human flk-1 coding sequence.

DE Human; receptor; protein; tyrosine kinase; pTK; primitive; mammalian;

KW hematopoietic cell; pHc; mature; mhc; fetal; liver kinase 2; flk-1;

KW liver; spleen; thymus; adult; brain; marrow; thymocyte; subset;

KW multipotential; T-lymphoid; lineage; stomach; kidney; lung; heart;

KW intestine; muscle; lymph node; ss.

XX Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 208..4311

FT /*tag= a

XX W09300349-A.

XX 07-JAN-1993.

XX 26-JUN-1992; 92WO-0505401.

XX 28-JUN-1991; 91US-0728913.

XX 15-NOV-1991; 91US-0793065.

XX 24-DEC-1991; 91US-0813593.

XX 02-APR-1992; 92WO-0502750.

XX	(UYPR-) UNIV PRINCETON.
PA	Lemischka IR;
XX	WPI: 1993-036323/04.
XX	P-PSDB; R31377.
DR	
XX	
XX	Nucleic acid encoding receptor protein tyrosine kinase - allows
PT	development of ligands to stimulate proliferation and/or
PT	differentiation of mammalian haematopoietic stem cells
XX	
PS	Claim 10; Fig 2; 78pp; English.
XX	
CC	This sequence encodes a human receptor protein tyrosine kinase which
CC	belongs to a new functional class of protein tyrosine kinases (ptks).
CC	However, this ptk is not in the same class as flk-2. ptk's in the same
CC	class as flk-2 are expressed in primitive mammalian hematopoietic
CC	(phc) cells but not in mature hematopoietic cells (mhc). This gene
CC	is expressed in more mature hematopoietic cells. The protein encoded
CC	by this sequence is an example of a receptor ptk and is called fetal
CC	liver kinase 1 (flk-1). flk-1 is expressed in fetal liver, spleen,
CC	thymus, brain, stomach, kidney, lung, heart and intestine, and adult
CC	brain, kidney, heart, spleen, lung, muscle, marrow and lymph nodes.
XX	
SQ	Sequence 5406 BP; 1412 A; 1299 C; 1422 G; 1273 T; 0 other;
Query Match 16.7%; Score 154.6; DB 14; Length 5406;	
Best Local Similarity 96.2%; Pred. No. 4.7e-34;	
Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;	
QY	718 ctgtgtcccgagcgggataaacctggctgacccgattccggcgagacccgtcgagccgc 777
Db	
1	ctgtgtcccgagcc-ggataacctggctgacccgattccggcgagacccgtcgagccgc 59
QY	778 ggctgagaccagggcgcggtgcgcccgctctccccggtcttgcgtcgggggcgcat 837
Db	
60	ggctgagaccagggcgcggtg-cccgcgctctccccggtcttgcgtcgggggcgcat 118
QY	838 accgcctctgactcttctggcgccagggagagagagagagagagagagag-aact 896
Db	
119	accgcctctgactcttctggcgccagggagagagagagagagagagagag-aact 178
QY	897 gggctctgtgtccca-gcgcgaggtgcagg 924
Db	
179	gggctctgtgtcccgagcgaggtgcagg 207
RESULT	9
Q40916	
ID	Q40916 standard; cDNA: 5406 BP.
XX	
AC	Q40916;
XX	
DT	19-OCT-1993 (first entry)
XX	
DE	Murine flk-1 cDNA.
XX	
KW	Murine; receptor; protein; tyrosine kinase; ptk; flk-1; flk-2; adult;
KW	family; primitive; hematopoietic cell; mature; fetal; liver; spleen;
KW	thymus; brain; stomach; kidney; lung; heart; intestine; bone marrow;
KW	muscle; lymph node; ss
XX	
OS	Mus musculus.
XX	
FH	Key Location/Qualifiers
FT	CDS 208..4311
FT	/*tag= a
FT	208..264
FT	/*tag= b
FT	/note= "Hydrophobic leader sequence"
FT	mat_peptide 265..4308

FT		/**tag= c
XX	WO9310136-A.	
FN		
XX		
PD	27-MAY-1993.	
XX		
PX		
PF	16-NOV-1992; 92WO-US09893.	
XX		
PR	15-NOV-1991; 91US-0793065.	
XX		
PA	(UYPR-) UNIV PRINCETON.	
XX		
PI	Lemischka IR;	
XX		
DR	WPI; 1993-182479/22.	
XX	P-PSDB; R37504.	
PT	Totipotent haematopoietic stem cell receptors, their ligands and	
PT	DNA sequences - for treating anaemia(s) and bone marrow damage	
PT	due to e.g. cancer chemotherapy or radiotherapy	
XX		
PS	Claim 14; Fig 2; 127pp; English.	
CC	This sequence encodes the murine receptor protein tyrosine kinase	
CC	(PTK), flk-1. This pTK is not in the same family as flk-2 (see also	
CC	Q40914-15) as it is expressed in primitive hematopoietic cells and	
CC	also in mature hematopoietic cells. flk-1 is expressed in fetal	
CC	liver, spleen, thymus, brain, stomach, kidney, lung, heart and	
CC	intestine and adult brain, bone marrow, kidney, heart, spleen, lung,	
CC	muscle and lymph nodes.	
XX		
SQ	Sequence 5406 BP; 1412 A; 1298 C; 1423 G; 1273 T; 0 other;	
Query Match	16.7%; Score 154.6; DB 14; Length 5406;	
Best Local Similarity	96.2%; Pred. NO. 4.7e-34;	
Matches 201; Conservative	0; Mismatches 4; Indels 4; Gaps	
QY	718 ctgtgtcccgagccggataacctggtgacccgattccggagcacaccgtgcagccgc	777
Db		
Db	1 ctgtgtcccgagcc-ggataacctggtgacccgattccggagcacaccgtgcagccgc	59
QY	778 ggcctggagccaggcggcggtgcccgcgctctccccggtcttgcgtcgggggcgcat	837
Db		
Db	60 ggctggagccaggcggcggtg-cccgctctccccggtcttgcgtcgggggcgcat	118
QY	838 accgcctctgtactcttttcggggccagggacggaggaaggagtctgtaccttgag-aact	896
Db		
Db	119 accgcctctgtactcttttcggggccagggacggaggaaggagtctgtacctgagaact	178
QY	897 gggctgtgtgccca-gcgcgaggtgcagg	924
Db		
Db	179 gggctgtgtgccagggcggaggtgcagg	207
RESULT	10	
Q81014		
ID	Q81014 standard; cDNA; 5406 BP.	
XX		
AC	Q81014;	
XX		
DT	18-AUG-1995 (first entry)	
XX		
DE	Flk1 receptor protein-tyrosine-kinase cDNA.	
XX		
KW	Mouse Flk1; receptor protein-tyrosine-kinase; hematopoietic stem	
KW	cell; fetal liver kinase; ds.	
OS	Mus musculus.	
XX		
Key	Location/Qualifiers	
CDS	208..4311	
FT	/*tag= a	

FT sig_peptide /product= Flk1 receptor protein-tyrosine-kinase
 FT 208..264 /*tag= b
 FT mat_peptide 265..4308 /*tag= c
 FT
 XX
 PN WO9500554-A.
 XX
 PD 05-JAN-1995.
 XX
 PF 17-JUN-1994; 94WO-US06944.
 XX
 PR 18-JUN-1993; 93US-0080244.
 PR 21-JUN-1993; 93US-0081508.
 PR 23-NOV-1993; 93US-0157490.
 XX
 PA (UYPR-) UNIV PRINCETON.
 XX
 XX Lemischka IR;
 XX
 DR WPI; 1995-052014/07.
 DR P-PSDB; R67817.
 XX
 XX Ligand for receptor protein tyrosine kinase - useful for the
 PT stimulation of primitive haematopoietic stem cells causing
 PT proliferation and/or differentiation
 PT
 XX Disclosure; Fig 2; 131pp; English.
 XX
 CC The sequence corresponds to a cDNA encoding a mouse Flk1 (fetal
 CC liver kinase) receptor protein-tyrosine-kinase. Flk1 is expressed
 CC in both primitive and mature hematopoietic cells, and in other
 CC tissues. The gene product is useful in isolation of receptor
 CC ligands, which have applications in diagnosis of bone marrow
 CC disorders and in stimulating proliferation and/or differentiation
 CC of primitive hematopoietic stem cells.
 XX
 SQ Sequence 5406 BP; 1412 A; 1298 C; 1423 G; 1273 T; 0 other;

Query Match 16.7% Score 154.6; DB 16; Length 5406;
 Best Local Similarity 96.2%; Pred. No. 4.7e-34;
 Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;
 Qy 718 ctgtgtccgcagcgggataacactggtgacccgattcccgagacacccgtgcagccgc 777
 ||||||||||| ||||||||||| ||||||||||| ||||||||||| |||||||||||
 Db 1 ctgtgtccgcagcc-ggataacctggtgacccgattcccgagacacccgtgcagccgc 59
 Qy 778 ggctggagccagggcgccggtgcccccgctctcccggttcttcgctgccccggcgcat 837
 ||||||||||| ||||||||||| ||||||||||| ||||||||||| |||||||||||
 Db 60 ggctggagccagggcgccggtg-ccccgctctcccggttcttcgctgccccggcgcat 118
 Qy 838 accgctctgtgactcttcttgcggccagggagagagagagagagagagagagagagagag 896
 ||||||||||| ||||||||||| ||||||||||| ||||||||||| |||||||||||
 Db 119 accgctctgtgactcttcttgcggccagggagagagagagagagagagagagagagagag 178
 Qy 897 gggtctgtgtccca-gcgcgaggtgcagg 924
 ||||||||||| ||||||||||| ||||||||||| ||||||||||| |||||||||||
 Db 179 gggtctgtgtcccgccgaggtgcagg 207

RESULT 11
 Q79070
 ID Q79070 standard; cDNA; 5406 BP.
 XX
 AC Q79070;
 XX
 DT 04-JUL-1995 (first entry)
 XX
 DE Mouse flk-1 cDNA.
 XX
 KW Fetal liver kinase-1; flk-1; protein tyrosine-kinase receptor;
 hematopoiesis; stem cell; ds.

XX Mus sp.
 OS
 XX
 FH Location/Qualifiers
 FT 208..4311
 FT /*tag= a
 FT sig_peptide 208..264
 FT /*tag= b
 FT mat_peptide 265..4308
 FT /*tag= c
 XX
 PN US5367057-A.
 XX
 PD 22-NOV-1994.
 XX
 PF 02-APR-1991; 91US-0679666.
 XX
 PR 02-APR-1991; 91US-0679666.
 PR 18-JUN-1991; 91US-0728913.
 PR 15-NOV-1991; 91US-0793065.
 PR 24-DEC-1991; 91US-0813593.
 PR 26-JUN-1992; 92US-0906397.
 PR 12-NOV-1992; 92US-0975049.
 PR 19-NOV-1992; 92US-0977451.
 PR 30-APR-1993; 93US-0055269.
 XX
 PA (UYPR-) UNIV PRINCETON.
 XX
 XX Lemischka IR;
 XX
 DR WPI; 1995-005894/01.
 DR P-PSDB; R67537.
 XX
 XX Murine flk-2 receptor protein tyrosine kinase - used to stimulate
 PT proliferation and/or stimulation of primitive mammalian
 PT haematopoietic stem cells in vitro or in vivo.
 XX
 XX Disclosure; Fig. 3A-3I; 69pp; English.
 XX
 CC cDNAs encoding receptor protein tyrosine-kinases, mouse fetal liver
 CC kinase-2 (flk-2), human flk-2 and mouse flk-1 are given in Q79068-70,
 CC respectively, and the deduced amino acid sequences in R67535-37,
 CC respectively.
 XX
 SQ Sequence 5406 BP; 1412 A; 1298 C; 1423 G; 1273 T; 0 other;

Query Match 16.7% Score 154.6; DB 16; Length 5406;
 Best Local Similarity 96.2%; Pred. No. 4.7e-34;
 Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;
 Qy 718 ctgtgtccgcagcgggataacactggtgacccgattcccgagacacccgtgcagccgc 777
 ||||||||||| ||||||||||| ||||||||||| ||||||||||| |||||||||||
 Db 1 ctgtgtccgcagcc-ggataacctggtgacccgattcccgagacacccgtgcagccgc 59
 Qy 778 ggctggagccagggcgccggtgcccccgctctcccggttcttcgctgccccggcgcat 837
 ||||||||||| ||||||||||| ||||||||||| ||||||||||| |||||||||||
 Db 60 ggctggagccagggcgccggtg-ccccgctctcccggttcttcgctgccccggcgcat 118
 Qy 838 accgctctgtgactcttcttgcggccagggagagagagagagagagagagagagagagag 896
 ||||||||||| ||||||||||| ||||||||||| ||||||||||| |||||||||||
 Db 119 accgctctgtgactcttcttgcggccagggagagagagagagagagagagagagagagag 178
 Qy 897 gggtctgtgtccca-gcgcgaggtgcagg 924
 ||||||||||| ||||||||||| ||||||||||| ||||||||||| |||||||||||
 Db 179 gggtctgtgtcccgccgaggtgcagg 207

RESULT 12
 T72119
 ID T72119 standard; cDNA; 5406 BP.
 XX
 AC T72119;

XX 19-AUG-1997 (first entry)
 XX Murine flk-1 receptor coding sequence.
 DE Human; fetal liver kinase 2; flk2; receptor protein tyrosine kinase;
 KW pTK; liver; spleen; thymus; adult; brain; bone marrow; primitive portion;
 KW haematopoietic hierarchy; extracellular domain; soluble form; ligand;
 KW proliferation; differentiation; mammalian; haematopoietic stem cell;
 KW macrocytic anaemia; aplastic anaemia; cancer; radiation; mouse; ss.
 XX OS Mus musculus.
 XX Key Location/Qualifiers
 FH CDS 208..431
 FT /*tag= a
 FT /product= Murine_flk-1
 XX US5621090-A.
 XX 15-APR-1997.
 XX 02-APR-1991; 91US-0679666.
 XX 26-JUN-1992; 92US-0906397.
 PR 02-APR-1991; 91US-0679666.
 PR 28-JUN-1991; 91US-0728913.
 PR 15-NOV-1991; 91US-0793065.
 PR 24-DEC-1991; 91US-0813593.
 XX (UYPR-) UNIV PRINCETON.
 PA Lemischka IR;
 PI WPI; 1997-235228/21.
 DR P-PSDB; W19875.
 XX Protein containing the extracellular domain of human flk-2 - used
 PT for identification of primitive haematopoietic cell proliferation
 PT and differentiation stimulatory ligands, e.g. for treating anaemia
 XX Disclosure; Fig 2; 55pp; English.
 XX This sequence encodes the murine fetal liver kinase 1 (flk-1). flk-1
 CC is a receptor protein tyrosine kinase (pTK). flk-1 is found in the
 CC same organs as flk-2 (see also T72117-18) as well as in fetal brain,
 CC stomach, kidney, lung, heart and intestine, and in adult kidney, heart,
 CC spleen, lung, muscle and lymph nodes. flk-1 is not a member of the
 CC same class of receptors as flk-2, as flk-1 may be found in more
 CC mature haematopoietic cells. The invention concerns a recombinant
 CC nucleic acid, preferably mRNA, which encodes a protein containing only
 CC the extracellular domain of human flk-2 and lacking the flk-2 intra-
 CC cellular catalytic domain. The resultant protein represents a soluble
 CC form of flk-2 which is used to isolate specific ligands for flk-2. These
 CC ligands can be used to stimulate proliferation and/or differentiation of
 CC mammalian haematopoietic stem cells, in vivo or in vitro, e.g. for
 CC treatment of macrocytic or aplastic anaemia or bone marrow damage caused
 CC by cancer treatment or radiation.
 XX Sequence 5406 BP; 1412 A; 1298 C; 1423 G; 1273 T; 0 other;
 SQ
 Query Match 16.7%; Score 154.6; DB 18; Length 5406;
 Best Local Similarity 96.2%; Pred. NO. 4.7e-34;
 Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;
 Qy 718 ctgttcccgagccgggataacctggctgaccgattccggagacacgctgcagcgcg 777
 Db 1 ctgttcccgagcc-ggataacctggctgaccgattccggagacacgctgcagcgcg 59
 Qy 778 ggtctgagccagggcggtgcccgcgctctcccggtcttgcgtgcggggcgcat 837
 Db 60 ggtctgagccagggcggtgcccgcgctctcccggtcttgcgtgcggggcgcat 118

QY 838 accgcctctgtacttcttgcggccagggacggagagagtgctgtgctgag-aact 896
 Db 119 accgcctctgtacttcttgcggccagggacggagagagtgctgtgctgagaaact 178
 QY 897 ggcctctgtgccca-gcgagaggtgcagg 924
 Db 179 gggctctgtgccca-gcgagaggtgcagg 207

RESULT 13
 T60765
 ID T60765 standard; DNA; 1267 BP.
 XX AC T60765;
 XX 15-OCT-1997 (first entry)
 XX Human KDR/flk-1 promoter bases (-780) to 487.
 XX KDR/flk-1; promoter; regulation; endothelial cell; transcription;
 KW gene therapy; treatment; specific; ds.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 FH primer_bind 1000..992
 FT /*tag= a
 FT misc_feature 781
 FT /*tag= b
 FT /note= "transcription start site"
 FT misc_feature 22..25
 FT /*tag= c
 FT /note= "potential cis-acting element (GATA)"
 FT misc_feature 190..195
 FT /*tag= d
 FT /note= "potential cis-acting element (E Box)"
 FT misc_feature 273..280
 FT /*tag= e
 FT /note= "potential cis-acting element (Ker)"
 FT misc_feature 506..611
 FT /*tag= f
 FT /note= "potential cis-acting element (E Box)"
 FT misc_feature 638..647
 FT /*tag= g
 FT /note= "potential cis-acting element (Ap2)"
 FT misc_feature 651..663
 FT /*tag= h
 FT /note= "potential cis-acting element (NFKB/Spl)"
 FT misc_feature 679..692
 FT /*tag= i
 FT /note= "potential cis-acting element (Spl/Ap2)"
 FT misc_feature 701..707
 FT /*tag= j
 FT /note= "potential cis-acting element (NFKB)"
 FT misc_feature 712..721
 FT /*tag= k
 FT /note= "potential cis-acting element (Ap2/Spl)"
 FT misc_feature 723..728
 FT /*tag= l
 FT /note= "potential cis-acting element (Spl)"
 FT misc_feature 737..742
 FT /*tag= m
 FT /note= "potential cis-acting element (Spl)"
 FT misc_feature 851..856
 FT /*tag= n
 FT /note= "potential cis-acting element (E Box)"
 FT misc_feature 887..890
 FT /*tag= o
 FT /note= "potential cis-acting element (GATA)"
 FT misc_feature 964..969
 FT /*tag= p
 FT /note= "potential cis-acting element (E Box)"

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 18:21:24 ; Search time 150.46 Seconds
(without alignments)
1137.715 Million cell updates/sec

Title: US-09-445-201-1_COPY_6036_6959

Perfect score: 924

Sequence: 1 gaagtcaacacccaagtgt.....tgcccagcgaggtgcagg 924

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents-NA:*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	169	18.3	5470	2	US-08-443-861-1
2	169	18.3	5470	4	US-08-193-829B-1
3	154.6	16.7	5406	1	US-07-813-593-3
4	154.6	16.7	5406	1	US-07-977-451-5
5	154.6	16.7	5406	1	US-07-946-507-3
6	154.6	16.7	5406	1	US-08-252-517-5
7	154.6	16.7	5406	1	US-07-906-397A-5
8	154.6	16.7	5406	2	US-08-601-891-5
9	154.6	16.7	5406	2	US-09-021-324-5
10	154.6	16.7	5406	5	PCT-US92-02750-7
11	154.6	16.7	5406	5	PCT-US92-05401-5
12	154.6	16.7	5406	5	PCT-US92-09893-5
13	70.6	7.6	2264	1	US-08-232-538-16
14	70.6	7.6	2264	2	US-08-786-164-16
15	40	4.3	1120	3	US-08-586-165-1
16	40	4.3	3937	3	US-08-586-165-8
17	40	4.3	7812	4	US-09-368-590-1
18	39	4.2	1026	1	US-07-975-526-6
19	37.4	4.0	1053	6	5352575-6
20	36	3.9	1641	1	US-08-385-229-1
21	36	3.9	1641	2	US-08-650-000-1
22	36	3.9	1641	6	5395760-1
23	36	3.9	2224	4	US-08-477-347-2
24	35.4	3.8	8438	1	US-07-945-283-1
25	34.6	3.7	2010	4	US-09-240-410-1
26	34.2	3.7	1028	4	US-08-118-200-1
27	34.2	3.7	1028	4	US-08-458-745-1

Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 29, Appli
Sequence 29, Appli
Sequence 1, Appli
Sequence 5, Appli
Sequence 17, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 14, Appli
Sequence 10, Appli
Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-08-443-861-1
; Sequence 1, Application US/08443861
; Patent No. 5851999
; GENERAL INFORMATION:
; APPLICANT: Ullrich, Axel
; APPLICANT: Risau, Werner
; APPLICANT: Millauer, Birgit
; APPLICANT: Gazit, Aviv
; APPLICANT: Levitzki, Alex
; TITLE OF INVENTION: Flk-1 Is A Receptor For Vascular
; TITLE OF INVENTION: Endothelial Growth Factor
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/443,861
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/193,829
; FILING DATE: 09-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7683-060
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)790-9090
; TELEFAX: (212)869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5470 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 286..4386

US-08-443-861-1

Query Match 18.3%; Score 169; DB 2; Length 5470;
Best Local Similarity 98.2%; Pred. No. 3.5e-36;
Matches 213; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 711 gccagactgtctcccgagccgggataaacctggtgacccgattccgcggacacccgctg 770
|||||
Db 70 GCCCAGACTGTGTCGCCGACGGGATAAACCTGCTGACCCGATTCCGCGGACACCGCTG 129
QY 771 -cagcccggtgagccagggcgccggtgcccccgccgtctccccggcttcttgccgtcg 829
|||||
Db 130 ACAGCCCGCGCTGGAGCAGCGCGCGGTGCCCGCGCTCTCCCGGCTCTTGGCGCTCGG 189
QY 830 gggcgatccgctctgtgactcttctggtggtgagcgaggaagaggtctgtgctt 889
|||||
Db 190 GGGC-CATACCGCTCTGTGACTTCTTTGGCGGCGGACGAGGAGGAGTCTGTGCTT 248
QY 890 gag-aactgggtctgtgcccc-gcgcgaggtgcag 924
|||||
Db 249 GAGAACTGGGCTCTGTGCCAGCGCGAGGTGCAGG 285

RESULT 2

US-08-193-829B-1
; Sequence 1, Application US/08193829B
; Patent No. 6177401
; GENERAL INFORMATION:
; APPLICANT: Ullrich, Axel
; APPLICANT: Risau, Werner
; APPLICANT: Millauer, Birgit
; APPLICANT: Gazit, Aviv
; APPLICANT: Levitzki, Alex
; TITLE OF INVENTION: Flk-1 Is A Receptor For Vascular
; TITLE OF INVENTION: Endothelial Growth Factor
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/193,829B
; FILING DATE: 09-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Cotuzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7683-060
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)790-9090
; TELEFAX: (212)869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5470 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 286..4386
US-08-193-829B-1

Query Match 18.3%; Score 169; DB 4; Length 5470;
Best Local Similarity 98.2%; Pred. No. 3.5e-36;
Matches 213; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 711 gccagactgtctcccgagccgggataaacctggtgacccgattccgcggacacccgctg 770
|||||
Db 70 GCCCAGACTGTGTCGCCGACGGGATAAACCTGCTGACCCGATTCCGCGGACACCGCTG 129
QY 771 -cagcccggtgagccagggcgccggtgcccccgccgtctccccggcttcttgccgtcg 829
|||||
Db 130 ACAGCCCGCGCTGGAGCAGCGCGCGGTGCCCGCGCTCTCCCGGCTCTTGGCGCTCGG 189
QY 830 gggcgatccgctctgtgactcttctggtggtgagcgaggaagaggtctgtgctt 889
|||||
Db 190 GGGC-CATACCGCTCTGTGACTTCTTTGGCGGCGGACGAGGAGGAGTCTGTGCTT 248
QY 890 gag-aactgggtctgtgcccc-gcgcgaggtgcag 924
|||||
Db 249 GAGAACTGGGCTCTGTGCCAGCGCGAGGTGCAGG 285

RESULT 3

US-07-813-593-3
; Sequence 3, Application US/07813593
; Patent No. 5185438
; GENERAL INFORMATION:
; APPLICANT: Lemischka, Ibor R.
; TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL
; TITLE OF INVENTION: RECEPTORS AND THEIR LIGANDS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCLONE SYSTEMS INCORPORATED
; STREET: 180 VARICK STREET
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/813,593
; FILING DATE: 19920415
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/793,065
; FILING DATE: 15-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/728,913
; FILING DATE: 28-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/679,666
; FILING DATE: 02-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Feit, Irving N.
; REGISTRATION NUMBER: 28,601
; REFERENCE/DOCKET NUMBER: LEM-3-PPP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-645-1405
; TELEFAX: 212-645-2054
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5406 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 208..4311

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/946,507
FILING DATE: 19920917
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/813,593
FILING DATE: 24-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/793,065
FILING DATE: 15-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/728,913
FILING DATE: 28-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/679,666
FILING DATE: 02-APR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Feit, Irving N.
REGISTRATION NUMBER: 28,601
REFERENCE/DOCKET NUMBER: LEM-3-PPP
TELEPHONE: 212-645-1405
TELEFAX: 212-645-2054
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5406 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 208..4311
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 208..4308
US-07-946-507-3

Query Match 16.7%; Score 154.6; DB 1; Length 5406;
Best Local Similarity 96.2%; Pred. No. 2.5e-32;
Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;
Qy 718 ctgtgtccgcagccggataaactgtgctgacccgattcccgagacacccgtgcagccgc 777
Db 1 CTGTGTCCTCCGAGCC-GGATAACCTGGCTGACCCGATTCCTCCGACACCCGTGCAGCCGC 59
Qy 778 gqctggagccagggcgccgggtgcccgcgctctccccggtcttgcgctgcccggcgccat 837
Db 60 GGCTGGAGCCAGCGCGCGGTG-CCTGCGCTCTCCCGGCTTGTGCGTGGCGGGCGCGAT 118
Qy 838 accgctctgtgactcttttgcggccagggacgaggaagagtgctgtgctgag-aact 896
Db 119 ACCGCCCTGTGTACTTCTTTTGCGGCCAGGACGGAGAGGAGTCTGTGCTGTGAGAACT 178
Qy 897 ggcctctgtgcca-gcgcaggtgacag 924
Db 179 GGGCTCTGTGCCAGCGCGGAGGTGCAGG 207

RESULT 6
US-08-252-517-5
Sequence 5, Application US/08252517
Patent No. 5548065
GENERAL INFORMATION:
APPLICANT: Lemischka, Ihor R.
TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL
TITLE OF INVENTION: RECEPTORS AND THEIR LIGANDS
NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:
ADDRESSEE: ImClone Systems Incorporated
STREET: 180 Varick Street
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10014
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/252,517
FILING DATE: 31-OCT-1994
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/977,451
FILING DATE: 19-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/906,397
FILING DATE: 26-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US92/05401
FILING DATE: 26-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: TW 81102961
FILING DATE: 15-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US92/02750
FILING DATE: 02-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/813,593
FILING DATE: 24-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/793,065
FILING DATE: 15-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/728,913
FILING DATE: 28-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/679,666
FILING DATE: 02-APR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Feit, Irving N.
REGISTRATION NUMBER: 28,601
REFERENCE/DOCKET NUMBER: LEM-3-7P
TELEPHONE: 212-645-1405
TELEFAX: 212-645-2054
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 5406 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
FEATURE:
NAME/KEY: CDS
LOCATION: 208..4311
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 265..4308
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 208..264
US-08-252-517-5

Query Match 16.7%; Score 154.6; DB 1; Length 5406;


```
Best Local Similarity 96.2%; Pred. No. 2.5e-32;
Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

QY 718 ctgtgtcccgagcgagcgagataaacctgctgacccgattccgagacaccgctgacgcgc 777
Db 1 CTGTGTCGCCGAGCG-CGATAAACCTGGCTGACCCGATTCGCGGACACCCGTCGACGCCG 59
QY 778 ggcctggagcagggcgccgctgccccgcgtctcccccggtcttcccggtcttgcgtgcggggcgcat 837
Db 60 GGCTGGAGCCAGGCGCGGTG-CGCCGCTCTCCCGGTCTTGGCTGCGGGGCCGAT 118
QY 838 accgcctctgtgacttcttgcgggcccagggagcagagagagctctgctgag-aact 896
Db 119 ACCGCCCTCTGTGACTTCTTTGCGGGCCAGGACGAGGAGTCTGTGCTTGAGAAACT 178
QY 897 ggcctctgtgccc-a-gcgcgaggtgcagg 924
Db 179 GGGCTCTGTGCCAGGCGCGAGGTGCAGG 207

RESULT 7
US-07-906-397A-5
; Sequence 5, Application US/07906397A
; Patent No. 5621090
; GENERAL INFORMATION:
; APPLICANT: Lemischka, Ihor R.
; TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: IMCLONE SYSTEMS INCORPORATED
; STREET: 180 VARICK STREET
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/906,397A
; FILING DATE: 19920626
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/813,593
; FILING DATE: 24-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/793,065
; FILING DATE: 15-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/728,913
; FILING DATE: 28-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/679,666
; FILING DATE: 02-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Feit, Irving N.
; REGISTRATION NUMBER: 28,601
; REFERENCE/DOCKET NUMBER: LEM-3-PPPPPP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-645-1405
; TELEFAX: 212-645-2054
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5406 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

; FRAGMENT TYPE: N-terminal
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 208..4311
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 208..4308
; US-07-906-397A-5

Query Match 16.7%; Score 154.6; DB 1; Length 5406;
Best Local Similarity 96.2%; Pred. No. 2.5e-32;
Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

QY 718 ctgtgtcccgagcgagcgagataaacctgctgacccgattccgagacaccgctgacgcgc 777
Db 1 CTGTGTCGCCGAGCG-CGATAAACCTGGCTGACCCGATTCGCGGACACCCGTCGACGCCG 59
QY 778 ggcctggagcagggcgccgctgccccgcgtctcccccggtcttcccggtcttgcgtgcggggcgcat 837
Db 60 GGCTGGAGCCAGGCGCGGTG-CGCCGCTCTCCCGGTCTTGGCTGCGGGGCCGAT 118
QY 838 accgcctctgtgacttcttgcgggcccagggagcagagagagctctgctgag-aact 896
Db 119 ACCGCCCTCTGTGACTTCTTTGCGGGCCAGGACGAGGAGTCTGTGCTTGAGAAACT 178
QY 897 ggcctctgtgccc-a-gcgcgaggtgcagg 924
Db 179 GGGCTCTGTGCCAGGCGCGAGGTGCAGG 207

RESULT 8
US-08-601-891-5
; Sequence 5, Application US/08601891
; Patent No. 5747651
; GENERAL INFORMATION:
; APPLICANT: Lemischka, Ihor R.
; TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ImClone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/601,891
; FILING DATE: 15-FEB-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/977,451
; FILING DATE: 19-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/906,397
; FILING DATE: 26-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US92/05401
; FILING DATE: 26-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: TW 81102961
; FILING DATE: 15-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US92/02750
; FILING DATE: 02-APR-1992
; APPLICATION NUMBER: US 07/813,593
```

STATE: New York
COUNTRY: U.S.A.
ZIP: 10014
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/021,324
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/977,451
FILING DATE: 1992-11-19
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/906,397
FILING DATE: 26-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US92/05401
FILING DATE: 26-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: TW 81102961
FILING DATE: 15-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US92/02750
FILING DATE: 02-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/813,593
FILING DATE: 24-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/793,065
FILING DATE: 15-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/728,913
FILING DATE: 28-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/679,666
FILING DATE: 02-APR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Feit, Irving N.
REGISTRATION NUMBER: 28,601
REFERENCE/DOCKET NUMBER: LEM-3-7P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-645-1405
TELEFAX: 212-645-2054
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 5406 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
FEATURE:
NAME/KEY: CDS
LOCATION: 208..4311
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 265..4308
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 208..264
US-08-601-891-5

Query Match 16.7%; Score 154.6; DB 1; Length 5406;

Best Local Similarity 96.2%; Pred. No. 2.5e-32;

Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

QY 718 ctgtgtcccgagccgggataaacttgctgacccgattccgaggaacaccgctgcagccgc 777

Db 1 CTGTGTCCCGCAGCC-GGATAAAGCTGGCTGACCGGATTCGCGGACACCCGCTGCAGCCGC 59

QY 778 ggcctggagccagggcgccgtgcccgcgcctctcccgcgtcttgcgctgcggggcgcat 837

Db 60 GCGTGGAGCCAGGCGCGCGGTG-CCCGCGCTCTCCCGGCTTTCGCGTGGGGGCGCAT 118

QY 838 accgcctctgtgactcttcttgcggccagggagcaggaagagtgctgtgctgag-aact 896

Db 119 ACCGCTCTGTGACTCTTTTTCGCGCCAGGACGAGGAGGAGTCTGTGCTGAGAAACT 178

QY 897 ggcctgtgtgcccga-gcgagagtgacgg 924

Db 179 GGGCTGTGTGCCAGGCGCGAGGTGCAGG 207

RESULT 9

US-09-021-324-5

Sequence 5, Application US/09021324

Patent No. 5912133

GENERAL INFORMATION:

APPLICANT: Lemischka, Ihor R.

TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL

TITLE OF INVENTION: RECEPTORS AND THEIR LIGANDS

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: ImClone Systems Incorporated

STREET: 180 Varick Street

CITY: New York

Query Match 16.7%; Score 154.6; DB 2; Length 5406;

Best Local Similarity 96.2%; Pred. No. 2.5e-32;

Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

QY 718 ctgtgtcccgagccgggataaacttgctgacccgattccgaggaacaccgctgcagccgc 777

Db 1 CTGTGTCCTCCGAGCC-GGATAAAGCTGGCTGACCGGATTCGCGGACACCCGTCGACCGC 59
QY 778 ggcctctgtgactcttcttgcggcgccggtgctctcccggtcttgcgctgctgctgag-aact 837
Db 60 GGCTGGAGCAGGCGCGGTG-CCCGCGTCTCTCCCGGTCTTGGCGTGGGGGGCGCAT 118
QY 838 accgctctgtgactcttcttgcggcgccggtgctctcccggtcttgcgctgctgag-aact 896
Db 119 ACCGCTCTGTGACTTCTTTTGGCGGGCAGGACGGAGGAGTCTGTGCTTGAGAACT 178
QY 897 ggcctctgtgactcttcttgcggcgccggtgctctcccggtcttgcgctgctgag-aact 924
Db 179 GGCTCTGTGCTCCGAGCGCGGAGGTGAGG 207

RESULT 10
PCT-US92-02750-7
; Sequence 7, Application PC/TUS9202750
; GENERAL INFORMATION:
; APPLICANT: LEMISCHKA, THOR R.
; TITLE OF INVENTION: Totipotent Hematopoietic Stem Cell
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: IMCLONE SYSTEMS INCORPORATED
; STREET: 180 VARICK STREET
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: US
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/02750
; FILING DATE: 19920402
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FEIT, IRVING N.
; REGISTRATION NUMBER: 28,601
; REFERENCE/DOCKET NUMBER: LEM-3-PPPT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-645-2054
; TELEFAX: 212-645-2054
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5406 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 208..4311
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 208..4308
PCT-US92-02750-7

Query Match 16.7%; Score 154.6; DB 5; Length 5406;
Best Local Similarity 96.2%; Pred. No. 2.5e-32;
Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;
QY 718 ctgtgtccgcagcgccggtgataacccggtgacccgattccgcggacacccgctgcagccg 777
Db 1 CTGTGTCCTCCGAGCC-GGATAAAGCTGGCTGACCGGATTCGCGGACACCCGTCGACCGC 59
QY 778 ggcctctgtgactcttcttgcggcgccggtgctctcccggtcttgcgctgctgag-aact 837
Db 119 ACCGCTCTGTGACTTCTTTTGGCGGGCAGGACGGAGGAGTCTGTGCTTGAGAACT 178
QY 897 ggcctctgtgactcttcttgcggcgccggtgctctcccggtcttgcgctgctgag-aact 924
Db 179 GGCTCTGTGCTCCGAGCGCGGAGGTGAGG 207

Db 60 GGCTGGAGCAGGCGCGGTG-CCCGCGTCTCTCCCGGTCTTGGCGTGGGGGGCGCAT 118
QY 838 accgctctgtgactcttcttgcggcgccggtgctctcccggtcttgcgctgctgag-aact 896
Db 119 ACCGCTCTGTGACTTCTTTTGGCGGGCAGGACGGAGGAGTCTGTGCTTGAGAACT 178
QY 897 ggcctctgtgactcttcttgcggcgccggtgctctcccggtcttgcgctgctgag-aact 924
Db 179 GGCTCTGTGCTCCGAGCGCGGAGGTGAGG 207
RESULT 11
PCT-US92-05401-5
; Sequence 5, Application PC/TUS9205401
; GENERAL INFORMATION:
; APPLICANT: Lemischka, Ibor R.
; TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL
; TITLE OF INVENTION: RECEPTORS AND THEIR LIGANDS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: IMCLONE SYSTEMS INCORPORATED
; STREET: 180 VARICK STREET
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/05401
; FILING DATE: 19920626
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Feit, Irving N.
; REGISTRATION NUMBER: 28,601
; REFERENCE/DOCKET NUMBER: LEM-3-PPPT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-645-1405
; TELEFAX: 212-645-2054
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5406 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 208..4311
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 208..4308
PCT-US92-05401-5

Query Match 16.7%; Score 154.6; DB 5; Length 5406;
Best Local Similarity 96.2%; Pred. No. 2.5e-32;
Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;
QY 718 ctgtgtccgcagcgccggtgataacccggtgacccgattccgcggacacccgctgcagccg 777
Db 1 CTGTGTCCTCCGAGCC-GGATAAAGCTGGCTGACCGGATTCGCGGACACCCGTCGACCGC 59
QY 778 ggcctctgtgactcttcttgcggcgccggtgctctcccggtcttgcgctgctgag-aact 837
Db 60 GGCTGGAGCAGGCGCGGTG-CCCGCGTCTCTCCCGGTCTTGGCGTGGGGGGCGCAT 118
QY 838 accgctctgtgactcttcttgcggcgccggtgctctcccggtcttgcgctgctgag-aact 896
Db 119 ACCGCTCTGTGACTTCTTTTGGCGGGCAGGACGGAGGAGTCTGTGCTTGAGAACT 178

QY 897 gggtctgtgcccc-gcgcgaggtgcagg 924
|||||
Db 179 GGCTCTGTGCCCCAGGCGGAGGTGCAGG 207

RESULT 12

PCT-US92-09893-5
; Sequence 5, Application PC/TUS9209893
; GENERAL INFORMATION:
; APPLICANT: Lemischka, Ihor R.
; TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL
; TITLE OF INVENTION: RECEPTORS AND THEIR LIGANDS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ImClone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/09893
; FILING DATE: 19921116
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Feit, Irving N.
; REFERENCE/DOCKET NUMBER: 28,601
; REGISTRATION NUMBER: LEM-3-7PT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-645-1405
; TELEFAX: 212-645-2054
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5406 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 208..4311
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 265..4308
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 208..264

PCT-US92-09893-5

Query Match 16.7%; Score 154.6; DB 5; Length 5406;
Best Local Similarity 96.2%; Pred. No. 2.5e-32;
Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;
QY 718 ctgtgtcccgagccggagataacctggctgacccgattccgcggagaccgctgcagccgc 777
|||||
Db 1 CTGTGTCGCGAGCC -GGATAACCTGCTGACCCGATTCGCGGACACCCGTCGAGCCGC 59
QY 778 ggcgtgagccagggcgccggtgccccgcgtctccccggtcttgcgtgcgggggcgcac 837
|||||
Db 60 GGCTGGAGCCAGGCGCGGTG -CCCCGCTCTCCCCGGTCTTTCGCGTGGCGGGGCGCAT 118
QY 838 accgctctgtgactcttgcgggcccagggaggaagagctctgtcctgaag-aact 896
|||||

Db 119 ACCGCTCTGTGACTTCTTTGCGGGCCAGGACGGAGAGAGAGTCTGTGCTCGAGAACT 178
QY 897 gggtctgtgcccc-gcgcgaggtgcagg 924
|||||
Db 179 GGCTCTGTGCCCCAGGCGGAGGTGCAGG 207

RESULT 13

US-08-232-538-16
; Sequence 16, Application US/08232538
; Patent No. 5712380
; GENERAL INFORMATION:
; APPLICANT: Thomas, Kenneth A.
; APPLICANT: Kendall, Richard L.
; TITLE OF INVENTION: INHIBITOR OF VASCULAR ENDOTHELIAL CELL
; TITLE OF INVENTION: GROWTH FACTOR
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: P.O. Box 2000 126 E Lincoln Avenue
; CITY: Rahway
; STATE: NJ
; COUNTRY: USA
; ZIP: 07065-0907
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,538
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Wallen, John W.III
; REGISTRATION NUMBER: 35,403
; REFERENCE/DOCKET NUMBER: 18888IA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-3905
; TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2264 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-232-538-16

Query Match 7.6%; Score 70.6; DB 1; Length 2264;
Best Local Similarity 63.9%; Pred. No. 6e-10;
Matches 140; Conservative 0; Mismatches 74; Indels 5; Gaps 2;

QY 711 gccagactgtgtcccgagccgggataacctggctgacccgattccgcggagaccgctg 770
|||||
Db 51 GCAGAAAGTCCGTCTGGCAGCCTGGATATCTCTCTACCGGACCGGACGCGCCCTG 110
QY 771 cagccgctgtgagccagggcgccggtgccccgcgtctccccggtcttgcgtgcggg 830
|||||
Db 111 CAGCGCGGTGCGGCGCGCGGCTCCCTAGCCCTGTGGCTCAACTGTCTGCGGTGCGGG 170
QY 831 g----gcgcataccgctctgtgactcttcttgcggggccagggagcgagagagctgtg 886
|||||
Db 171 GTGCGCGGAGTTCCACCTCGCGGCTCTTCTCTAGACAGCGGCTGGGAGAAAGACCGG 230
QY 887 cctgagaact-gggctgtgtcccgagcgaggtgcagg 924
|||||
Db 231 CTCCCGAGTTCGCGCATTTCCCGCGGCTCGAGGTGCAGG 269

RESULT 14
US-08-786-164-16

```

: Sequence 16, Application US/08786164
: Patent No. 5861484
: GENERAL INFORMATION:
: APPLICANT: THOMAS, KENNETH A.
: APPLICANT: KENDALL, RICHARD L.
: TITLE OF INVENTION: INHIBITOR OF VASCULAR ENDOTHELIAL
: TITLE OF INVENTION: CELL GROWTH FACTOR
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Merck & Co., Inc.
: STREET: P.O. Box 2000, 126 E. Lincoln Ave.
: CITY: Rahway
: STATE: NJ
: COUNTRY: USA
: ZIP: 07065-0900
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: Microsoft Word 6.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/786,164
: FILING DATE: 21-JAN-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Mark Hand, J
: REGISTRATION NUMBER: 36,545
: REFERENCE/DOCKET NUMBER: 18888DA
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 908-594-3905
: TELEFAX: 908-594-4720
: TELEX:
: INFORMATION FOR SEQ ID NO: 16:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2264 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
: US-08-786-164-16

```

Query Match 7.6%; Score 70.6; DB 2; Length 2264;
Best Local Similarity 63.9%; Pred. No. 6e-10;
Matches 140; Conservative 0; Mismatches 74; Indels 5

Qy	711	gccagactgtccgcgacgcgggataacctgctgctgacccgattccgcggacacccgtg	770
Db	51	GCAGAAAGTCGGCTGGCAGCCCTGGGATATCCTCTCTACGGCACCGCAGACGCCCTG	110
Qy	771	cagccgcgctgaagccagggcccggtgccccgcgcctctcccgcgtttgcgctcgagg	830
Db	111	CAGCCGCGTGGGCCCGGGCTCCCTAGCCCTGTGGCTCAACTGTCTCGCTCGGG	170
Qy	831	g----gcgcataccgcctgtgactcttttgcggccaggagcggagagagctctgtg	886
Db	171	GTGCGCGAGTTCACCTCGCGCCTCCCTCTCTAGACAGGCGCTGGGAGAAAGAACCG	230
Qy	887	ccctgagaact-gggctctgtgccagcgcgaggtgcagg	924
Db	231	CTCCCAGATTCCGGCAATTTGCGCGGCTCGAGTGCAGG	269

RESULT 15

RESULTS
US-08-586-165-1/c
; Sequence 1, Application US/08586165
; Patent No. 6054298
; GENERAL INFORMATION:
; APPLICANT: Laufer, Edward M.
; APPLICANT: Orozco, Olivia E.

```

1 APPLICANT: Tabin, Clifford J.
2
3 TITLE OF INVENTION: Fringe Proteins and Pattern Formation
4
5 NUMBER OF SEQUENCES: 9
6
7 CORRESPONDENCE ADDRESS:
8
9 ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
10
11 STREET: Two Militia Drive
12
13 CITY: Lexington
14
15 STATE: Massachusetts
16
17 COUNTRY: US
18
19 ZIP: 02173
20
21 COMPUTER READABLE FORM:
22
23 MEDIUM TYPE: Floppy disk
24
25 COMPUTER: IBM PC compatible
26
27 OPERATING SYSTEM: PC-DOS/MS-DOS
28
29 SOFTWARE: PatentIn Release #1.0, Version #1.30
30
31 CURRENT APPLICATION DATA:
32
33 APPLICATION NUMBER: US/08/586,165
34
35 FILING DATE: 16-JAN-1996
36
37 CLASSIFICATION: 800
38
39 ATTORNEY/AGENT INFORMATION:
40
41 NAME: Granahan, Patricia
42
43 REGISTRATION NUMBER: 32,227
44
45 REFERENCE/DOCKET NUMBER: H095-05
46
47 TELECOMMUNICATION INFORMATION:
48
49 TELEPHONE: (617) 861-6240
50
51 TELEFAX: (617) 861-9540
52
53 INFORMATION FOR SEQ ID NO: 1:
54
55 SEQUENCE CHARACTERISTICS:
56
57 LENGTH: 1120 base pairs
58
59 TYPE: nucleic acid
60
61 STRANDEDNESS: single
62
63 TOPOLOGY: linear
64
65 FEATURE:
66
67 NAME/KEY: CDS
68
69 LOCATION: join(1..240, 244..474, 478..531, 538..579, 58
70
71 LOCATION: .678, 682..687, 691..807, 811..843, 850..870
72
73 LOCATION: 874..990, 994..1056, 1060..1083, 1087..1104,
74
75 LOCATION: .1119)
76
77 US-08-586-165-1

```

Query Match 4.3%; Score 40; DB 3; Length 1120;
Best Local Similarity 51.7%; Pred. No. 0.076;

	Matches	91;	Conservative	0;	Mismatches	85;	Indels	0;	Gaps	0;
QY	748	accgattccgcggacacccgtcgcagccgcggctggagccagggcgcgcggtgcgcccgcc	gcgc	807						
Db	248	ACGGGTACGGGACGACCCCGCGCCCGCGCGGTCCCAGCGCGCGCGGGTCCC		189						
QY	808	tctcccgcgtcttgctgcgtggggggcgcataccgcctctgtgaattcttttcgggccagg		867						
Db	188	CGGACAGCTGGCTCCCGCGGGCGCGCGCTCCCGCTTCGGGGAGGGCGCTGCTGGCGCGG		129						
QY	868	gacggagaagagttctgtgctctgagaaactgggtctctgtgccacgcgcgagatgcag		923						
Db	128	CGGCGCGCGGCGCGCGCGGGCGCGCGGGGGGCTGTCCCGGGGGCGACGACAGC		73						

Search completed: June 21, 2001, 18:21:46
Job time: 17499 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 15:34:03 ; Search time 4579.66 Seconds
(without alignments)
1907.224 Million cell updates/sec

Title: US-09-445-201-l_COPY_6036_6959
Perfect score: 924
Sequence: 1 gaagttcacaccgaatgt.....tgccagcgaggtgcagg 924

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_estl1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: gb_est41.*
42: gb_est42.*
43: gb_est43.*
44: gb_est44.*
45: gb_est45.*
46: gb_est46.*
47: gb_est47.*

44: em_esthum10.*
45: em_esthum11.*
46: em_esthum12.*
47: em_esthum13.*
48: em_esthum14.*
49: em_esthum15.*
50: em_esthum16.*
51: em_esthum17.*
52: em_esthum18.*
53: em_esthum19.*
54: em_esthum20.*
55: em_esthum21.*
56: em_esthum22.*
57: em_esthum23.*
58: em_esthum24.*
59: em_esthum25.*
60: em_esthum26.*
61: em_esthum27.*
62: em_esthum28.*
63: em_estin1.*
64: em_estin2.*
65: em_estin3.*
66: em_estin4.*
67: em_estin5.*
68: em_estomi.*
69: em_estom2.*
70: em_estov1.*
71: em_estov2.*
72: em_estpl1.*
73: em_estpl2.*
74: em_estpl3.*
75: em_estpl4.*
76: em_estpl5.*
77: em_estpl6.*
78: em_estpl7.*
79: em_estpl8.*
80: em_estpl9.*
81: em_estpl10.*
82: em_estro1.*
83: em_estro2.*
84: em_estro3.*
85: em_estro4.*
86: em_estro5.*
87: em_estro6.*
88: em_estro7.*
89: em_estro8.*
90: em_estro9.*
91: em_estro10.*
92: em_estro11.*
93: em_estro12.*
94: em_estro13.*
95: em_estro14.*
96: em_estro15.*
97: em_estro16.*
98: em_estro17.*
99: em_estro18.*
100: em_estro19.*
101: em_estro20.*
102: gb_est25.*
103: gb_est26.*
104: gb_est27.*
105: gb_est28.*
106: gb_est29.*
107: gb_est30.*
108: gb_est31.*
109: gb_est32.*
110: gb_est33.*
111: gb_est34.*
112: gb_est35.*
113: gb_est36.*
114: gb_est37.*
115: gb_est38.*
116: gb_est39.*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est167:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*

```

```

190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

TITLE JOURNAL COMMENT

RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp
URL: http://genome.rtc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermotabilization and thermoactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsuina,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.

FEATURES source

Location/Qualifiers
1..192
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="CS30002A01"
/clone_lib="RIKEN full-length enriched, 12 days embryo
spinal cord"
/tissue_type="spinal cord"
/dev_stage="12 days embryo"
/lab_host="DHI0B"
/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGATCCAGTTCGAGTTAAATATCCGCCGCCGCC 3']. cDNA
was prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGATTCGAGTTAAATATCCGCCGCCGCC 3']. cDNA
was cleaved with XhoI and BamHI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT ORIGIN

Query Match 17.0%; Score 157.4; DB 161; Length 192;
Best Local Similarity 89.0%; Pred. No. 2.6e-33;
Matches 170; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
Qy 640 gtctcagaccacccagagaaagctgtttctcttagatcgccgagccgtaccgcg 699
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2 GTCTCTAGGAACCCCAAGAAAGTAGGTGTGTTCTTATAGTCGCGGAACCGTACCCCG 61
Qy 700 caggactaaagccagactgtgtcccgagcgggataacctggctgaccgattccgc 759
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 CAGGAATGAAGCCAGATGTGTCCCGAACCGGAAACCTGTTGAACCGATTCCCG 121
Qy 760 ggcacacgctgcagccgcggtgagccagggcgccggtgcgccgcgtctcccggtct 819
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 122 GGAACCCCTGCAGCGGGGGTGGAGGAGGAGGGCGGTGACCCGCCGCCCTCCCGGTCT 181
Qy 820 tgcgctgcggg 830

Db 182 CGCTCTGTGGG 192

RESULT 3

BF543874 207 bp mRNA EST 11-DEC-2000
UI-R-C3-tg-c-12-0-UI.r1 UI-R-C3 Rattus norvegicus cDNA clone
UI-R-C3-tg-c-12-0-UI 5', mRNA sequence.
BF543874
ACCESSION BF543874
VERSION BF543874.1 GI:11634981
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 207)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
cDNA Library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.G.E. Consortium at
LLNL (info@image.llnl.gov). IMAGE ID= 1785489
Seq primer: M13 Forward.

FEATURES source

Location/Qualifiers
1..207
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C3-tg-c-12-0-UI"
/clone_lib="UI-R-C3"
/dev_stage="adult"
/lab_host="DHI0B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C3
library is a subtracted library of a series, ultimately
derived from a mixture of individually tagged normalized
libraries from rat placenta, adult lung, brain, liver,
kidney, heart, spleen, ovary, muscle, and 8, 12 and 18-day
embryos, after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: UI-R-C3, UI-R-C2p, UI-R-C1
, UI-R-C0, UI-R-A1, UI-R-E1. The tag is a string of 3-5
nucleotides present between the Not I site and the
oligo-dT track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (UI-R-C3) was constructed as follows: PCR amplified
cDNA inserts from UI-R-C2p clones from which 3' ESTs had
been derived was used as a driver in a hybridization with
the UI-R-C2p library in the form of single-stranded
circles. The remaining single-stranded circles (subtracted
library) was purified by hydroxyapatite column
chromatography, converted to double-stranded circles and
electroporated into DHI0B bacteria (Life Technologies) to
generate the UI-R-C3 library. This procedure has been
previously described (Bonaldo, Lennon and Soares, Genome
Research 6:791-806, 1996)".
Research 6:791-806, 1996

BASE COUNT ORIGIN

36 a 58 c 64 g 39 t

```

Query Match      16.4%; Score 151.2; DB 150; Length 207;
Best Local Similarity 88.0%; Pred. No. 1.4e-31;
Matches 176; Conservative 0; Mismatches 23; Indels 1; Gaps 1;

Qy 654 aagagagtaagctgtgttcttagatcgccggaccgctaccggcaggactgaaagcc 713
Db 8 AAGAGAGTAGGCTGTCTCTCTAGATCGCCGCGAAGCGTACCCGCGACGACTGAAAGCC 67

Qy 714 cagactgtgtcccgacgggataaactggctgctaccgattccggcgagaccctgcag 773
Db 68 CAGATTGTGCTCTACGCGGGATAAAGCTGGCTGACCGGTGTCCGCGACACTGCTGCAC 127

Qy 774 ccgcggtgagcagggcgcc-ggtgccccgcgctctcccgctctgtgcgctcggggg 832
Db 128 CCGCGGTGAGCAGCGCGCGGTCCTCCGCTTCCAGGTTTGACTGCAGGAG 187

Qy 833 cgcataccgctctgtgact 852
Db 188 CGCGTAACGCTTCTGTGACT 207

RESULT 4
BB593727 193 bp mRNA EST 30-NOV-2000
LOCUS BB593727 RIKEN full-length enriched, 4 days neonate male adipose
DEFINITION Mus musculus cDNA clone B430204G23 5', mRNA sequence.
ACCESSION BB593727
VERSION BB593727.1 GI:11490329
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T., Hodoiyama,Y.,
Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J., Kojima,Y., Konno
H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K.,
Nomura,K., Numazaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C.,
Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shinagawa,A.,
Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka
T., Toya,T., Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K.,
Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
Unpublished (2000)
CONTACT: Yoshinide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome.res@rtc.riken.go.jp,
URL:http://genome.rtc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermozabillation and thermoactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
Location/Qualifiers
1..193
/organism="Mus musculus"

```

```

/db_xref="taxon:10090"
/clone="B430204G23"
/clone_lib="RIKEN full-length enriched, 4 days neonate
male adipose"
/sex="male"
/tissue_type="adipose"
/dev_stage="4 days neonate"
/lab_host="DH10B"
/notes="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGAGAGAGGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 229.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGATTCGAGTTAATTAATATCCGCCGCCGCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."
BASE COUNT 38 a 44 c 82 g 29 t
ORIGIN

Query Match      15.8%; Score 146.2; DB 161; Length 193;
Best Local Similarity 85.3%; Pred. No. 3.5e-30;
Matches 163; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Qy 638 gagtcctcagaccccaagagtaagctgtgttcttagatcgccggaccctacc 697
Db 3 GAGTCTCTCAGAACCCAGAGAGTAAAGTGTTCTTACATCGGGGGAACGATACC 62

Qy 698 ggcaggactgaagcccgactgtgtcccgagcgagataacctggctgacccgattcc 757
Db 63 GCGAGGAGTGAAGCGCAGAGATGTCTCCGAGGCGGGAATCTGGCTGAACGATCC 122

Qy 758 gcggacaccctgcagccgcgctggagccagggcgccgctgcccgcgctctcccg 817
Db 123 GGGGAAACCGTTGACGCGGGGTGGAGGAGCGCGCGGTGCCCGCGGTCTCCCGGT 182

Qy 818 ctgctgcgtgcg 828
Db 183 CTGGGGGTGGG 193

RESULT 5
AL589555 335 bp mRNA EST 07-MAR-2001
LOCUS DKFP451M0219_r1_451 (synonym: hlccl) spinal cord Homo sapiens cDNA
DEFINITION clone DKFP451M0219 5', mRNA sequence.
ACCESSION AL589555
VERSION AL589555.1 GI:13243327
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.
Bloeker,H., Boecher,M., Brandt,P., Mewes,W., Well,B. and Wiemann
S.
EST (Bloeker,H., Boecher,M., Brandt,P., Mewes,H.W., Well,B. and Wiemann
Wiemann,S.)
JOURNAL Unpublished (1999)
COMMENT Contact: Bloecker H
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

```



```

DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR16H04 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL050923
VERSION     AL050923.1 GI:4932977
KEYWORDS    fruit fly.
SOURCE      Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 692)
AUTHORS     Direct Submission
TITLE       Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mamoosier in the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES             Location/Qualifiers
     source           1..692
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR16H04"
                     /note="end : TET3"

BASE COUNT  118 a 69 c 104 t 333 others

Query Match      5.0%; Score 46.4; DB 219; Length 692;
Best Local Similarity 17.4%; Pred. No. 0.033;
Matches 56; Conservative 139; Mismatches 126; Indels 0; Gaps 0;

Qy 425 ccaccagaaacccagctgctccagattgtctcagatgcgacttgccgcgcgcgcacag 484
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 370 SNVAGVRVRMVVVGVSVRVSACACSMCACGCCSSASVCASRMSSGSCSGSRSR 429

Qy 485 ttccgggttagtgaggagtggtggtggaacccggaacccaaacccatggttccagtg 544
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 430 SRSWGSGRVGGSGVSGSGSSAGRVMAASGAGAGSGVSGSSGBCACVSVMS 489

Qy 545 ggggggtggtggcagcagcaggttcccccacccctcccgtaagtacccccccccattc 604
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 490 SSRVAVSMGSGVCGCVVVMVCMCRVRMMKCCCGVCGVSSGACCCCVMAACRCVAV 549

Qy 605 gctagtgtagcggcgctctcttttgcctgagctcagcagcccaagagagtaag 664
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 550 SCRAGAACMAAASVMSVCVVAVSVVMCMASMASMSGAVSVCMRSMASVAVG 609

Qy 665 ctgtgtttcttagtcgcgcgcgcgtaccgccgagactgaagcccgactgtgtc 724
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 610 CACASRMVACCMSCAMASVSGSGARGAVRSASSSSASASMSMCMCAVAMSCVASV 669

Qy 725 ccgcagccgggataacctggc 745
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 670 SRSSVAVGVSVMMSVSAVC 690

RESULT 14

```

```

CNS012PK      1101 bp      DNA      GSS      26-JUL-1999
Drosophila melanogaster genome survey sequence T7 end of BAC
BACN08K01 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION  AL101858
VERSION     AL101858.1 GI:5613469
KEYWORDS    fruit fly.
SOURCE      Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 1101)
AUTHORS     Direct Submission
TITLE       Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBelOBAC11.

FEATURES             Location/Qualifiers
     source           1..1101
                     /organism="Drosophila melanogaster"
                     /plasmid="pBelOBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN08K01"
                     /note="end : T7"

BASE COUNT  80 a 251 c 335 g 63 t 372 others

Query Match      5.0%; Score 46.2; DB 219; Length 1101;
Best Local Similarity 34.8%; Pred. No. 0.041;
Matches 81; Conservative 1; Mismatches 151; Indels 0; Gaps 0;

Qy 471 gcccccgccgacagttccgggttagtggtggagtggtggcgtagggagaaacccggaaccccaaa 530
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 402 GCTCCCGCNCACGAGGAGGNGCNCNGGGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG 461

Qy 531 cctggtatccagtggtggcggtgacgacgagagagagagagagagagagagagagagagagag 590
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 462 GNGCNCGACGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 521

Qy 591 ccccgcccccattcgctagtgtagcggtgctctcttcttgcctgagctcagtagc 650
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 522 GGCCCGCNCNNGGCGNNGNNGNNGCNCNGCGGCGCNCNNGNNGNNGNNGNNGNNGNNGC 581

Qy 651 cccaagagagtaagctgtgttcttccttagatgcgcgagaccgaccccgagcag 703
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 582 NGGCGNANNNGCGCNCNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGC 634

RESULT 15
CNS006XX/c
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL066051
VERSION     AL066051.1 GI:4945019
KEYWORDS    fruit fly.
SOURCE      Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

```


Search completed: June 21, 2001, 15:34:07
Job time: 7651 sec

